

USING A NEURAL NETWORK TO
UPDATE AN EXPERT SYSTEM

By

MATTHEW R. THOMPSON

Bachelor of Science

Oklahoma State University


Stillwater, Oklahoma

1996

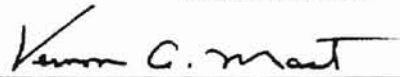
Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
May, 1998

USING A NEURAL NETWORK TO
UPDATE AN EXPERT SYSTEM


Thesis Approved:



Thesis Adviser







Dean of The Graduate College

ACKNOWLEDGMENTS

I wish to express my appreciation to my major advisor, Dr. McTernan for his supervision, guidance, inspiration, and friendship. I would also like to thank Dr. McTernan for helping me with advice on the starting of my career. I would also like to thank the other members of my committee Dr. Mast and Dr. Sanders for their help and assistance throughout this research project. Special thanks go out to Kevin Howell and Jami Striegel for their help obtaining some information that was used throughout the thesis.

I would like to express my sincere gratitude to my parents, Larry and Celia Thompson, for keeping me focused on my goals and following my dreams. I would also like to thank them for their encouragement, love and support.

I wish to also thank Coach Pat Jones who through his efforts help me through school with a full ride scholarship in football. I would also like to thank him for allowing me to continue playing the game that I love.

Finally, I would like to Thank Dr. Hughes and the rest of the Department of Civil and Environmental Engineering for putting up with my antics for so many years.

TABLE OF CONTENTS

Chapter	Page
1.0 INTRODUCTION.....	1
1.1 Congress's Approach to Cleaning up the Environment	2
1.2 Using Expert Systems to Evaluate Environmental Problems.....	4
1.3 Using CORA as the Modeled Expert System.....	4
1.4 Definition of the Scope of the Project	6
2.0 NEURAL NETWORKS.....	7
2.1 History of the Neural Network	7
2.2 How the Neural Network Works	10
2.3 Neural Network Parameters.....	15
2.4 Types of Neural networks.....	18
3.0 EXPERT SYSTEMS.....	21
3.1 History of the Expert System	21
3.2 How Expert Systems Operate.....	22
3.3 CORA as an Expert System	24
4.0 MATERIALS AND METHODS	27
4.1 Neural Network Setup for Containment.....	28

Chapter	Page
4.2 Inputting the Spreadsheet into the Neural Network	36
4.3 Neural Network Parameters	38
4.4 Training The Network	44
4.5 Spreadsheet Modifications for Treatment Options	42
4.6 Updating Costs of CORA's Remediation Technologies	43
5.0 RESULTS AND DISCUSSION	46
5.1 Model Dimensions	46
5.2 Training the Network	48
5.3 Cost Results	51
5.4 Examples to Compare the Neural Network to CORA	53
6.0 SUMMARY	72
REFERENCES	75

LIST OF FIGURES

Figure	Page
1. Biological Neural Network: A Brain	10
2. Artificial Neural Network: A Computer	11
3. Cost Analysis from CORA	26
4. CORA Output Example	30
5. Containment Flow Chart.....	33
6. Sample Portion of Spreadsheet.....	File Folder @ End
7. Screen from Neuralyst used for Loading the Spreadsheet.....	37
8. Neural Network Parameters	39
9. Treatment Flow Chart.....	44
10. Entire Spreadsheet used to Train the Neural Network	File Folder @ End
11. Complete Flow Chart Showing the Decision Tree in CORA	47
12. Example Problem - Case 1 CORA Results.....	55
13. Example Problem - Case 2 CORA Results.....	58
14. Example Problem - Case 3 CORA Results.....	61
15. Example Problem - Case 4 CORA Results.....	64
16. Example Problem - Case 5 CORA Results.....	67
17. Neural Network results for the Example Problem	69

LIST OF TABLES

Table	Page
1. Containment Questions used by CORA	29
2. Training and Test Cases from the Neural Network	34
3. Treatment Questions used by CORA.....	42
4. Remediation Schemes used in Containment and Treatment.....	49
5. Results of the Training Runs.....	50
6. Update Costs for CORA	52
7. Comparison of Results for the Example	71

1.0 Introduction

The industrial revolution that occurred during the nineteenth and twentieth centuries changed the United States. It caused almost as many problems as it solved. The invention of complex machines that bolstered the revolution was actually the reason for the improvements in society. The onset of these machines created situations that the new revolution could not foresee. They required fuel, lubricants, cleaners, degreasers, and other industrial solvents and keep the revolution moving. These products come from crude oil and other fluids (mainly fresh water and salt water) drawn from the ground and brought to the surface through pipes made of various components. These fluids from the ground were then stored in tanks until processed or used at the industries. These early day transportation and storage methods were not very efficient at transporting the fluids from the drilling process to the point of use. The pipes used for transportation often leaked. The early storage containers were made of redwood that would leak until the wood swelled from the moisture. The drilling rigs would blow out, spraying oil out of the top of the dike until the well pressure could be stabilized. Maintaining the machines required cleaning and maintenance, after which the various fluids would be simply discarded. This often meant that the fluids were thrown down the drain, poured out on the ground or discharged into a stream without any type of treatment.

These early industrial practices have led to many types of environmental problems associated with the ground, water, and air found in North America (Masters, 1991). The leaks and spills caused by the fuel and lubricants of the industrial revolution were not perceived at the time to cause significant problems. The dumping or discarding of the

fluids often created an aesthetics problem with the neighboring people. As time passed, these environmental problems caused by the dumping started to mount. The types of problems that arose were undrinkable water supplies, sickness, and even death (Masters, 1991). These problems were often not directly associated with the industrial practice until much later in the century. The problem only seemed to rise due to increasing numbers of industrial facilities and more potent chemicals and escalated well into the late 1970's until the federal government stepped in and developed new laws.

1.1 Congress's Approach to Cleaning up the Environment

Congress began applying minor regulatory pressure towards industry in the late 1950's with an environmental policy that dealt with clean water. Although it was not the first environmental law in the United States, the Rivers and Harbors Act dealing with dredging in the 1890's was the first environmental law (Hughes, 1995). The 1950's clean water law was written under the common law system. This means that the property owner retained control of environmental conditions, unless it could be proven in a court of law that a particular incident harmed someone off of the property. In 1969, Congress passed the National Environmental Policy Act (NEPA) making the first time that the United States defined its environmental objective (Sullivan, 1997). This law addressed environmental problems without the traditional common law approach. This act did help alleviate some of the problems in certain cases, but did not solve all the environmental problems. NEPA only applied to federal government controlled land or persons dealing directly with the federal government. That meant that environmental situations on private

lands were not controlled under this policy. The next government action came in the early 1970's following two major environmental disasters. One was an oil spill off the coast of California that coated the Santa Barbara beaches with oil, and the Cuyahoga River in Ohio caught fire. The Environmental Protection Agency (EPA) was created from the environmental departments of several agencies to provide a unified focus to the nation's environmental effort. Congress also passed several laws to rectify past environmental problems. These acts included the Clean Water Act of 1972, which dealt with discharges into surface waters, and the Research Conservation and Recovery Act (RCRA) of 1976, which addressed solid waste (Thomas, 1995). The acts that were created in the late 1970's were not effective until the early 1980's. The law mitigated past or abandoned hazardous waste sites was the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA). Each act and law was an attempt to strengthen and/or eliminate weak or missing areas that dealt with environmental pollution and control. Congress has also passed laws in the 1980's and 1990's to assist in the environmental protection plan of the United States. They include the Clean Air Act (CAA), Safe Drinking Water Act (SDWA), Superfund and Reauthorization Act (SARA), Toxic Substances Control Act (TOSCA), and the Pollution Prevention Act (PPA) to name a few (Findley, 1991).

These laws have many things in common, they state that the contaminated areas or events must be dealt with and cleaned to appropriate levels to alleviate any harm to the public. The appropriate levels are not always clear. The problem is how do we reach these clean-up levels? The laws are not specific how to obtain the final concentration

levels of the contaminated area. As a result, engineering designs used to remediate a contaminated site relies on accepted current practices when remediating a particular site.

1.2 Using Expert Systems to Evaluate Environmental Problems

The processes that were used in the early stages were decision trees based on limited knowledge and experience. Different situations required following different branches of the tree. This method was only as good as the engineers practical experience and the questions in the decision tree. The early engineers' decision tree was greatly effected by the type of information that the decision tree contained. If the engineer had only a limited knowledge of information of a particular type of remediation, then all applicable remediation schemes may not be considered. As a result, the best remediation scheme for a particular problem may not be evaluated.

The engineering world approached this problem by creating decision trees that were based on the most up-to-date technology. The easiest way to get this up-to-date technology was through a computerized decision tree model that was updated with the most current innovations in science and technology. These newly created models were called expert systems. By proceeding through the program and imputing information required by the expert system on a certain site, it will identify remediation schemes applicable to the unique situations that exist at the site.

1.3 Using CORA as the Modeled Expert System

The expert system that was used during this research project was called Cost of Remedial Action (CORA) (CH2M Hill, 1990). CORA is a program that was created and

maintained for the Environmental Protection Agency until the early 1990's. CORA at the time of its last update in 1990 gave the user a way to evaluate the site in a screening matrix of remedial activities by asking specific questions about the site. The questions ranged from what type of chemicals were present, soil types, and information concerning the hydrology of the site. The program then interprets each answer and recommends several applicable remediation schemes. The user then inputs the total amount of contaminated materials and the recommended level of clean-up for that site. This information results in CORA giving the user an approximate cost of the remediation and clean-up. The costing module includes everything from clearing and grubbing to moving off-site. CORA needs to be updated with regards to costs, new remediation technologies, and program default parameters.

CORA can be updated by rewriting and revising the computer code. As an alternative, this research applied a neural network model. A neural network works by pattern recognition that comes from training the system to learn. Training means that the neural network is capable of looking at the inputs and outputs of a given situation to predict the results of different situation. The structure of the neural network is modeled after the human brain. In general, the brain works by connecting millions of neurons with synapses (Figure 1, in Chapter 2, shows an illustration of the brain deals with thoughts in the human brain). All the neurons in a brain are connected to each other to allow the brain to process many types of complex information easily. The computer counterpart can achieve similar results with increased efficiency and speed relative to conventional computer codes. Neural networks have the capacity of learning the complicated decision tree associated with CORA or other expert systems while being able to update the

underlying data base and decision algorithm. The advantage of updating CORA by using a neural network is that the time needed to update the system was relatively minimal and a knowledge of CORA's programming language is not needed. Similarly, the information contained in CORA was updated for greater applicability. Once the training is complete, the updates are very fast, usually less than five minutes. The only computer information that the user must know to run the neural network is a basic knowledge of computer spreadsheets augmented with the neural network "add-on". In this manner, a tool was developed quickly and efficiently which built upon previous engineering knowledge with readily available skills and software.

1.4 Definition of the Scope for the Project

The scope of this project was confined to only sites that contain Volatile Organic Compounds (VOCs) as the chemical of concern. Although, CORA has the capability of giving remediating alternatives for many other categories of chemicals that might exist at a site. VOCs were chosen for two reasons. The first was that CORA was a very complex program that was designed to handle every condition that a user might see at a job site. The second reason was that VOCs are a common contaminant problem that exist in Oklahoma as a direct result of the oil industry and other industries. The neural network was designed and set-up with the same questions as the CORA program for all types of situations where VOCs may need to be remediated. The ultimate goal of this research project is to achieve a tool that is based on the slightly dated information from CORA with the updated information that has been added through the neural network that can be

used to help make decisions in future environmental remediation sites. The neural network that was used in this research project was Neuralyst (Shih, 1994). This tool will also have the capability of being updated with any type of information through the use of the neural network.

2.0 Neural Networks

2.1 History of Neural Networks

Neural networks have a history that dates back to the early 1940's. The first mention of a neural network was in an article by McCulloch and Pitts in 1943 (Wilson and Sharda, 1992). The McCulloch-Pitts paper started the examination of how the human brain functions. McCulloch-Pitts believed that the brain exists as a series of interacting parts, which evolve continuously, whose sole function is dependent upon the connection of these intermediate parts (Wilson and Sharda, 1992). This paper was given credit for starting the research for processing systems of data based on the function of the human brain. As a result, this research was the first ever done on the neural network.

Donald Hebb, a psychologist, in 1949 further advanced the science of the human brain with a new theory (Wilson and Sharda, 1992) that stated that the brain operates by a collection of neurons called assemblies. These neurons process the brain's impulses and convert them into an appropriate behavior. Hebb also believed that the neurons were only interconnected by what he termed self-organization. He believed that the interconnections between neurons helped strengthen the pathways.

In 1957, a scientist named Rosenblatt made the next major stride in the neural network field by developing one of the first learning systems (William and Sharda, 1992). The system was a neuron-like device that was called the "Perceptron." This was the first neuron type program or machine that could learn from data input.

The biggest break in the early history of the neural network occurred in 1960 when two electrical engineers, Windrow and Hoff, described the first neuron based

computer (Wilson and Sharda, 1992). This neuron based computer is the foundation for the modern neural network training of computer programs.

An opposite view was published that produced evidence that questioned the earlier work. This damaging view was published in a book called "Perceptrons" co-authored by Minsky and Papert in 1969 (Wilson and Sharda, 1992). The opinion was that the neuron based computer (Windrow and Hoff) did and would not work. They disputed the Rosenblatt's perceptron device and stated that the device could not solve the simplest of problems. They further stated that the neuron device, while having attractive features, could not be carried over to the multi-layer networks proposed by the earlier scientist (Wilson and Sharda, 1992). This sent ripples through the scientific community that seriously disputed the earlier claims of Rosenblatt. A lack of technology further fueled this argument. The computers in the 1960's and 1970's were severely limited in speed and capacity. They were large but very slow by today's standards. As such they could not process problems of the complexity and sophistication posed by typical neural network approaches. Federal research money was temporarily directed away from neural network topics, resulting in a thirteen year period where little or no additional development occurred.

In 1982, John Hopfield restarted neural network research and restored its credibility to the scientific community (Wilson and Sharda, 1992). His work proved that neural networks were able to solve various types of problems from simple to complex by applying a learning system program that was trained to learn.

In 1986, Rumelhart, Hinton, and Williams derived a learning algorithm for Rosenblatt's Perceptron device. The algorithm was developed from the work performed

by Rosenblatt as well as Windrow and Hoff in the late fifties and early sixties. This learning algorithm was termed a “back propagation” network. Today, “Back propagation is the most widely used training algorithm for multi-layer networks today” (Wilson and Sharda, 1992). In the end, the perseverance of these scientists and engineers has given the technical community a very innovative way to solve modern complex problems.

2.2 How The Neural Network Works

Neural networks operate in a very complex way that is hard to completely understand. The important thing for environmental engineering applications is that to use the program only requires knowledge of the program and its features and not the underlying mathematics. These networks have been proposed for many applications with various types of problems. On the simplest level, a neural network is setup like a human brain with many simple elements (neurons) that work together in parallel. The following two figures show how the brain functions in parallel and how the neural network attempts to simulate the human brain. Figure 1 shows that the neurons in the human brain are setup in parallel and how they are connected to each other. A more complex look at the neurons would show multiple layers of these neurons that would be interconnected.

Biological

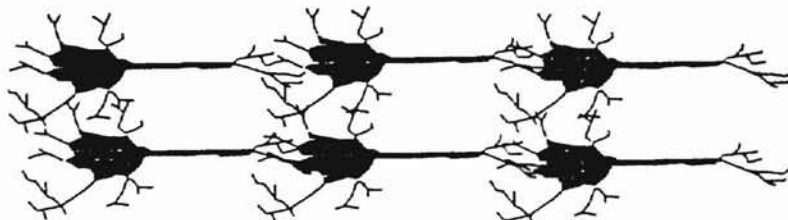


Figure 1: A Biological Neural Network: A Brain (Powers, 1994)

Figure 2 illustrates how the computer based structure of neural networks. The neurons that make-up each layer are connected to each other to create a network of neurons.

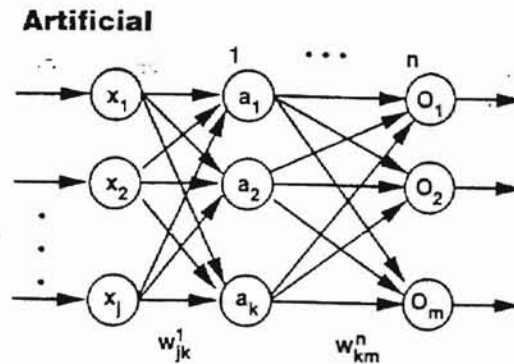


Figure 2: Artificial Neural Network: A Computer (Powers, 1994)

The “X’s” in the figure represent the input layer of a neural network. The information is then spread throughout the network through the hidden layers which are represented with the “A’s”. After the information is processed in the hidden layers the information goes into the output layer defined as “O’s”. The difference between the biological process of the brain and the artificial processes associated with a computer must be dealt with to produce an affective artificial neural network.

Neural networks can have one or more layers of neurons. These networks normally allow the user to have a range of 2 - 6 layers. The layers are separated into three different categories. The first layer of neurons in the network is termed the input layer. Input neurons are used to define the problem that is to be solved. The last layer of neurons is the output layer. Output neurons are used by the computer for the results of the neural network. All other neuron layers between the input and out are called hidden layers. Hidden layers are used by the neural network to process the inputs while predicting the outputs. These layers are responsible for the all the number crunching and pattern association that is needed to convert or model the input information into the

desired output. The layers can be interconnected together, or only connected on each individual layer. The connections between all neurons carry weights that contain the knowledge of the system. The neurons then sum the obtained information from other surrounding neurons or external input stimuli (Wilson and Sharda, 1992). This information is then transferred by each neuron in a non-linear fashion to ultimately achieve a trained system. The neurons with the summed information will then internally process the information and distribute the information to the next appropriate neuron or to the external output. The exact way in which the neural network operates is based on the structure of the computer code.

A decision has to be made on how many layers are needed for every neural network. The higher the number of hidden layers the more generalizations can be made in the network. The larger number of layers also allows the program to use fewer numbers of neurons in the development of the neural network. It is reported that most neural networks can be solved with three layer systems (Shih, 1994). The neural network program used in this research project has a default size of three layers. System processing speed may be able to be increased with the addition of more neurons and/or hidden layers. In some cases, too few layers and neurons in the neural network can cause the network to stall and never reach the desired goal of 100% pattern recognition. Pattern recognition occurs when the neural network produces the same results as the target results that were inputted into the network by the user. The reason being that too few neurons force the network to place a large amount of information into an individual neuron resulting in a loss of efficiency. In the case of too many neurons, there is so little information in an individual neuron it is almost a wasted space. This in turn will result in

a slowing down of the network. Lowering the number of neurons will result in a speeding up of the network. A problem with a larger layer system is the potential for a great loss of time in the training period. The larger systems take longer to train due to the size of the neural network. They have more layers and more neurons. The smallest neural network that can efficiently give satisfactory predictable results are typically the best networks as far as time is concerned. Selecting the number of layers and number of neurons is an iterative process.

A neural network can solve both linear and non-linear problems. As with any “learning” experience the network is initially very prone to mistakes. Once calibrated, the neural network can become very precise. The network is ultimately driven by the type of problem to be solved. The neural network begins as a collection of rules and inputs that must be taught to achieve better results. This learning process is termed training.

A neural network can be trained by either supervised or unsupervised training (Wilson and Sharda, 1992). Supervised training requires the user to describe both the problem domain and the answer (desired output). This allows the neural network to learn that this particular input will result in this output. A control group is required to teach the network the appropriate learning response from the training. The control group that contains data with known inputs and outputs is used as a calibration instrument for the network. After the input is entered, the neural network processes the inputs and the given output until the system can train itself to produce the desired output. Training starts when the network is given a problem and higher control neurons try to apply the hidden neurons to generate the expected answer. The specific learning algorithm that is used

during the training will determine how the neuron's interconnections are weighed. The algorithm will then correct the weights of the neurons due to the differences in the actual and desired outputs. The network is continually updating the system until the desired outputs are achieved. In other words, after the calibration is completed the system will adjust the weighted values of the neurons so that when a given input is received, the correct output will be calculated (Gifford, Gifford, and Gifford, 1997).

In an unsupervised learning system, there is not a target answer to which the network tries to learn. This pattern tries to learn by the process of repeated exposures. Wilson and Sharda state that, “. . . this kind of learning can be envisioned as the neural network appropriately self-organized or clustering its neurons related to the specific desired task” (1992).

A properly functioning neural network will be able to learn the underlying problem from a unique input signal and be able to map it onto a continuous output curve. This process is called generalization (Gifford, Gifford, and Gifford, 1997). The number of hidden neurons in a network is very important to the quality of the answer. If a network does not have enough hidden neurons, the neurons will overgeneralize the problem. The network will sort too much information in not enough categories. As a result, the answer from the network will not reflect the complexity of the problem resulting in a wrong answer. Networks with too many hidden neurons produce trivial answers because the network has too many places to store information.

The neural network has many advantages over conventional models which have been used in the past. Traditional models develop formulae that mimic reality. The data for a particular situation is used to fit the model. A neural network develops functional

relationships between the input and outputs. Neural networks can be adapted to the most complex problems where other models are often too simplified. The training process teaches the program the importance of every neuron instead of focusing on a single point such as a maximum or minimum data point. Each neuron is assigned a significant (weight) and is identified by the classification of the connection. This allows the program to establish useful relationships between the neurons in the network. The network can be as complex as the original problem with no worry about whether a human can understand the processing of the information. The computer will give a result that an engineer or scientist should be able to understand.

2.3 Neural Network Parameters

There are several parameters that are used by a neural network to help train the system. They include the learning rate, momentum, training tolerance, and testing tolerance. The process used to determine these parameters is an iterative one. Each parameter will have to be evaluated to determine the correct number to enhance the neural network's performances.


The learning rate is used to control the way in which the error is used to correct the weights in the neural network for each training case. This correction in the error is the way the network trains itself. The learning rates do this by reducing problems dealing with stability. The range of the learning rate is from 0 to 1. The higher the learning rate the more problems that can arise in the stability of the neural network. The lower number will reduce unstable behavior. Unstable behavior is when the neural network plateaus

during training (number of correct answers does not increase as time progresses). The lower the learning rate the longer the network will take to develop. A learning rate that is too low will result in a neural network that can not be trained without a significant amount of computer time. If the network can be trained without the extra time required by a low learning rate, then the low learning rate is not needed. The same results can be produced without the extra computer time.

Momentum deals with the amount of previous error that is applied to the weight adjustment in each training case. If a user defines momentum to be 0.5, then the weight adjustment will be 50% from the current error and 50% of the adjustment will be applied to the previous case within the neuron. The neuron then takes the starting value given by the user and using an exponential decay, reduces the amount of error associated with the next neuron by 50% each time. The higher the momentum is set the smoother the training process will operate. It also allows the neural network to adjust for any unusual circumstances that might arise in the training process. A setting of 1 would cause the network to use 100% of the previous error for the weight adjustment. This means that for the first training case the error from the previous example, there is no previous example because it is the first training case, would have its weights adjusted by 0%. The next training case would then take 100% of the error associated with the previous case (0) and adjust it by the 0% again. This will continue no matter how many training cases. Therefore, the training will never stop because the network is unable to train itself by adjusting the error. This means that the momentum must be lower than 1 or 100% (Shih, 1994). A lower momentum can be used if the data appears to be regular and smooth in appearance. The lower rate will result in longer training times.

Training tolerance is used during the actual training procedure of the neural network. It tells the neural network how much training is needed to consider itself trained. In other words, this number reflects how close the network must come to the desired answer to consider that particular output as correct and trained. Tighter training tolerances force the computer to come closer to the desired targets. For instance, if the training tolerance is set at 0.2, then the computer will have to come within $\pm 20\%$ of the target answers to be considered trained. It basically increases the accuracy of the network by decreasing the difference between the targets and the network generated outputs. This increased accuracy will result in longer and rougher running during the training times, but might train the network a little more than necessary to get the desired results. If the training tolerance is set too low the network may encounter a phenomenon called overtraining. This occurs when the network is so concerned predicting the answer that it loses its ability to generalize the training data. This means that the training of the network will never stop because the machine can not achieve the proper answers that would stop the training procedure.

Testing tolerance indicates the same thing as training tolerance except that it is used by the computer after the training period when predicting the test cases and tells the network how close its generated values must be to the target answers to be considered a correct response. The training tolerance allows the user to define to the network how close the output of the test cases must be to the target answer given into the training case. For example, if the user gives a value of 0.4, if the computer's output is within 40% of the answer it is considered correct. Testing tolerances can be set up to 0.5 or 50% for true- false type neural networks. The reason being that true is represented by a one where

a false identification is represented by a zero. This feature will have to be set closer to zero for networks that are using number values instead of true false. 

2.4 Types of Neural Networks

There are many types of neural networks used to solve problems. Three of the most commonly used neural networks are multi-layer feed forward networks, Hopfield networks, and Kohonen's feature maps (Wilson and Sharda, 1992). The multi-layer feed forward networks have many applications in classification and forecasting problems.

This network type consists of layers of neurons which pass information in one direction from input layers, through hidden layers, and finally to output layers of the network. The neurons of each layer are only connected to the subsequent layer of neurons. The hidden layers are allowed to develop their own internal representation of mapping input to output. Overall, the network operates with a non-linear behavior, which allows the neurons of the neural network model to learn many different types of input-output relationships. The training of the system is done by a back propagation training algorithm which is a type of supervised training. OR/MS Today defines back propagation as an “. . . attempt to minimize an error measure such as the sum of squared error during the training process” (1992). Back propagation is used when the error is determined and a part of it is propagated back through the network within the neurons. The error is used to adjust the weights and threshold values at each neuron, so that the next time, the error within the network response will be less for the same inputs and outputs. Back propagation is applied continuously until the network error in the results

fall below the training or testing rate set by the user (Shih, 1994). Back propagation is one of the most common training algorithms used today.

The second type of neural network which is commonly used is a Hopfield Network. This model is based on a series of papers presented by John Hopfield in the 1980's (Wilson and Sharda, 1992). It is composed of highly interconnected networks of non-linear neurons. It is a single layer of neurons in which every neuron is connected to every other neuron. The system's output from a neuron depend on the previous values stored in the neuron. Interconnections between the neurons are weighted, which eventually will lead to a feasible solution. The system creates an energy function, which represents the goal of the neural network by stopping the evolution of the network when the right output is calculated (Wilson and Sharda, 1992). The Hopfield network starts with neurons, and assigns a weight to each of the interconnections. The neurons and their weights are adjusted over time as the network runs. Eventually the neurons stabilize as they approach the final answer as determined by the energy function. This network does not use any type of training to teach this particular system. The interconnections between the neurons and the weight they carried tend to be fixed once established. This type of network is best used for an optimization problem, especially when the application can take advantage of the parallelism of the network (Wilson and Sharda, 1992).

The last major type of neural networks is the Kohonen's self-organization network. This network uses an unsupervised approach to teach the network. The network is based on the premise that the human brain classifies some of its information based upon an applied external stimulus. The algorithm used in Kohonen's networks form what he called a "feature map" (Wilson and Sharda, 1992). This is where the

neighborhoods are sensitive to the same certain types of inputs. This network is most often used in a cluster analysis problem. The network has also been shown that it can handle certain types of organizational problems.

3.0 Expert Systems

Expert systems are one of the most commonly used computer tools for making decisions. The field of expert systems is at the forefront of commercialization in computer science (Walker and Miller, 1986). Expert systems are defined as a “. . . computer system (hardware and software) that simulates human experts in a given area of specialization” (Castillo, Gutierrez, and Hadi, 1997). The reason expert systems are becoming popular in the 1990’s is because of the very high level of expertise required to address increasingly complex problems. These expert systems, if designed properly, contain an extremely high level of human expertise required to make potentially life threatening decisions. This in turn, gives a high level of expertise to everybody who uses the expert system appropriately and responsibly.

3.1 History of Expert Systems

Expert systems fall under a new field called artificial intelligence. The official beginning of artificial intelligence was at a convention at Dartmouth College in 1956 (Walker and Miller, 1986). A prediction made at this conference was that within the next 25 years everyone would be involved in recreational activities, while their computers back at the office would do all the work (Walker and Miller, 1986).

The development of artificial intelligence was slowed during the 1960’s and 1970’s when programs that became more difficult and complex than originally thought were installed on computers lacking power and speed. Through these lean years, research continued and successful artificial intelligent programs or systems were developed in the

early 1980's. The major advancement that spurred this new growth in artificial intelligence was the rapid growth and power of the modern computer. As a result, there are numerous types of artificial intelligent programs in use in commercial businesses and industry. The majority of the artificial intelligence and expert systems are being used in the business world, but increasingly there are significant environmental examples.

3.2 How Expert Systems Operate

Expert systems are based directly on knowledge created or derived from their human counterparts. This process is called knowledge engineering (Williams, 1986).

Knowledge engineering involves 4 components. The knowledge engineering must:

- (1) define the expert system domain,
- (2) elicit desired information from human experts,
- (3) incorporate that knowledge into acceptable form for the knowledge base, then
- (4) test the system to evaluate its robustness and accuracy (Williams, 1986).

The knowledge base contains facts, rules, and relationships derived from expert experience relating to specific situations. Once the knowledge base has been established the computer-based expert system must simulate the thinking process of a human expert. The knowledge base must then use an inference engine to apply the knowledge base efficiently. The inference engine automatically identifies and evaluates any knowledge from the knowledge base that would be applicable to each given situation. This process of going through the knowledge base and the inference engine is not always easy. The knowledge engineer must incorporate not only general rules about a situation, but also the

specialized rules associated with special circumstances. The information can be added later when the error is detected. The problem is that the problem may not be caught until the system has been sent to the public. For this reason, expert systems have become iterative processes of multiple refinement. (Williams, 1986)

Expert systems offer several advantages to a company that uses them. The expert system allows the design engineer to make decisions with the most current knowledge as long as the computer program is updated. If the expert system is not updated continuously, the system can still be very useful. The updates normally involve minor changes in the decision tree or the addition of new processes. As a result, the updated version will provide only a small change in the output. If the system is not updated over a period of time then the small changes will add up to make your system out of date. Expert systems are capable of outperforming their human counterparts that become bored, forgetful, or developing tunnel vision (Williams, 1986). The speed of the decision making process used in expert systems via the computer is far superior to a human. The limitation is that the computer is only as smart as the human who created the program. Expert systems are also susceptible to any wrong information that is inputted into the expert system. Human experts have two qualities that supersede the computer. Humans have considerable knowledge in specific areas and can have effective strategies for quickly sorting through knowledge when faced with problems (Williams, 1986). Clearly the use of an expert system in association with a human expert can improve the decision making process.

3.3 CORA as an Expert System

Cost of Remedial Action (CORA) was the expert system that was simulated with a neural network in this research project. CORA was originally written for the Environmental Protection Agency (EPA) to allow the user to define a site remediation approach and to assign associated costs (1990). The program is intended for users who are knowledgeable about the site. The relatively little information that is required concerning the site includes hydraulic conductivities or soil types and limited concentration range information (CH2M Hill, 1990). It is assumed in the program that the remediation process is in the earliest stages of the remedial investigation. From the information provided by the user, CORA will then recommend several relevant remediation actions for the site as well as estimating costs. It is to be used specifically at sites that have not had a feasibility study.

The program is comprised of two independent components: an expert system and a cost system. The expert system component requires the user to respond to questions about various contamination types and levels as well as remedial action goals and conditions present at a site. The expert system will recommend to the user a range of remediation actions that can be considered at the site. CORA's expert system functions as the knowledgeable advisor and creates the knowledge base. The system then evaluates the answers to the questions to determine what additional facts are required for that particular site. The expertise used by CORA for the inference engine originates from EPA policy considerations and technical feasibility (CH2M Hill, 1990). CORA makes a

decision on different remediation schemes by looking at the site characteristics and chemicals of concern. The program does not evaluate any alternatives based on cost.

After the user evaluates the options created in the expert system component, the user then inputs the desired remediation actions into the cost module of the program. The different types of remediation actions that were selected are then economically evaluated separately. An example of an economic evaluation for a remediation scheme is located on the following page in Figure 3, representing a general cost estimate generated by CORA for an excavation example. This figure also shows the type of questions that are required for the cost evaluation for a particular remediation scheme. Both the estimated capital and O&M costs are presented.

Within the cost module, remediation schemes will be costed out after a few more questions about the site are answered. These additional questions cover topics such as the area of contamination, type of contamination, and level of protection required for the public and workers. This is accomplished by CORA with the use of default values in certain places throughout the cost modules (CH2M Hill, 1990). This feature allows the user to establish a price before all the site information is collected. Since the economic evaluation can be based on default values, the program requires the user to input a level of confidence. The program does allow the user to input more specific site information. Costs generated by the program cover capital cost and the first year of operational costs (CH2M Hill, 1990). CORA does not provide all the economic elements that may exist in all remedial actions. For example, cost of equipment, decontamination of buildings, and/or applicable permits are not included in the overall costs provided by CORA.

***** VERSION 3.0 DRAFT *****

DATE: 03/29/98
TIME: 17:39:34

CORA SOIL EXCAVATION COST MODULE (201)

SITE NAME: EXAMPLE FOR THESIS
 OPERABLE UNIT: ENTIRE SITE ESTIMATED START: MID FY 1998
 SCENARIO: MODULES COMMON TO ALL SCENARIOS
 RUN BY: Matt Thompson PHONE NUMBER:

INPUTS		RESULTS	
Parameter	Value	Component	Total
TEST		BYPRODUCTS FOR TRANSPORT/DISPOSAL:	
Soil type	4		
Depth of excavation (ft)	5	DRUMS	0
1. Steel sheeting or		CONTAMINATED SOIL (CY)	
2. side slope?	1	(SWELL FACTOR=1.45)	128
Horizontal component	1		
Length of excavation (ft)	25		
Width of excavation (ft)	25		
Depth of cover above			
contaminated materials (ft)	2		
Depth of contaminated excav.			
w/o continuous sampling (ft)	0		
Depth of contaminated excav.			
w/continuous sampling (ft)	3		
Thickness of lifts (inches)	12		
Number of drums	0		
Pct. of contaminated zone	0		
Base air monitoring required?	N		
Pct. of backfill available			
onsite	75		
INPUTS		RESULTS	
Parameter	Value	Component	Total
Protection level for:		COST FOR ALL EXCAVATIONS	
Uncontaminated materials	A		
Contaminated materials	A	CAPITAL COST	240,000
Temperature (degrees F)	65	O & M COSTS	0
Confidence level	H		

*** Excavation depth cannot exceed 25 feet. For excavations deeper than 25 feet, complex site-specific sheeting, bracing, dewatering, terracing and haul roads may be required. Excavation for depths deeper than 25 feet should be scoped and costed on a site-specific basis.

Figure 3: Cost Analysis from CORA (CH2M Hill, 1990)

4.0 Materials and Methods

The ultimate goal of the project was to have a 100% recognition rate in an efficient manner. The 100% recognition rate occurs once the training of the network leads to a 100% prediction of the test cases. The neural network will tell the user that all variables are correct while no variables are incorrect. When this criteria is met, the goals of the neural network have been accomplished and the network is ready for almost any type of containment or treatment scenarios dealing with VOC's previously addressed by CORA.

The first step of the research project was to evaluate CORA and decide the exact scope of this project for the neural network. This was required to help develop the spreadsheet that will be used in the neural network. A spreadsheet is needed as a place to set up the information from CORA. The spreadsheet will then be used to input the information into the neural network. The scope was fixed on a potential site that could exist in the state of Oklahoma. Consistent with input to CORA, the soil condition of the site was limited to homogeneous contaminated unsaturated soils. The neural network was set up to evaluate both containment and treatment remediation schemes while the type of chemicals of concern for the potential site were limited to Volatile Organic Carbons (VOC). The types of soils that were evaluated were clay and silt. It was felt that these represented most of the potential site characteristics for contaminated site in Oklahoma.

4.1 Neural Network Setup for Containment

The neural network setup was adversely impacted because the amount of data needed to achieve a 100% pattern recognition rate from the neural network from CORA was unknown. The overall size of the spreadsheet based model will not be finalized until the training and testing of the neural network is finished. It is an iterative process requiring many trials. The research had to solve this dilemma by going back into CORA and collecting data. These data were used to train the neural network (described in Section 2.2). The questions that CORA asked, except for the information described in Section 4.0, were true-false questions. CORA as well as the neural network gave different remediation schemes depending on the answers to these questions.

The network was established by looking at the types of questions and answers that were asked by CORA. The key to this project was to directly relate CORA to the spreadsheet, which in turn was used to load and train the network. The neural network was based on the use of a spreadsheet. The size of the spreadsheet was not known at the start of the research project. That is, the different site scenarios would lead to different questions from CORA. The spreadsheet must be large enough to encompass every type of question for treating or containing VOC's. The actual structure of the network required that questions that were related to a certain issue be incorporated into the network together. Different issues would then need to be related to the network as a different entity. The actual network could not be setup until the data from CORA were collected and input into the spreadsheet.

The first data collected from CORA dealt with different scenarios associated with containment schemes. An example of an output from CORA for a typical containment scenario, is located on the following two pages in Figure 4. The first part of Figure 4 shows the type of questions with answers that were asked by CORA for one containment scenario. The second part of Figure 4 shows the remediation schemes that were recommended based on these question. An entire list of all the questions that are possible within the containment section are listed below in Table 1.

Table 1: CORA Containment Questions*

- What waste types apply to the site (Q1)
- What response action do you wish to consider (Q2)
- What types of contaminants are in the soil (Q4)
- Will excavation of the contaminants cause environmental or public impacts (Q5)
- Is the contaminated soil a hazardous waste (Q6)
- Is the contaminated soil concentration above land disposal restrictions (Q7)
- Is an onsite landfill reasonable (Q8)
- Select all types of contaminants in leachate from landfill (Q8-a)
- Are contaminated soil located in a 100-year flood plain (Q8-b)
- Is a shallow aquifer present that would not allow a below grade landfill (Q9)
- Type of discharge option either water reinjection (Q10)-water infiltration (Q10-a)-discharge to POTW (Q-10b)-discharge to surface water (Q10-c)
- Could site conditions threaten health or safety of unauthorized visitors (Q11)
- Are exposed soils on the site exposed to erosion (Q12)
- Pick the location of the site: above floodplain (Q12-a)-at base of hill above floodplain (Q12-b)-in floodplain (Q12-c)

*The questions are not sequentially numbered due to other questions that resulted in the treatment section of CORA.

The data were collected by picking true and false responses for the questions listed above.

In some cases all of these questions would not arise. For instance, if question 6 was answered false and question 7 was answered true then questions 8 - 10c would not be asked. The next question would be 11 (is the site considered a hazard to unauthorized personnel). There are several other instances that involve these if-then questions in this

***** VERSION 3.0 DRAFT *****

DATE: 03/19/98
TIME: 11:16:30

CORA EXPERT SYSTEM

RUN:

RUN BY: Matt Thompson

SITE: Thesis Example

CONTAMINATED AREA: Example

WASTE TYPE: HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

INPUT

Response type: Containment

Soil contaminant: Volatile organic compounds

Excavation acceptable: True

Material in question is hazardous: True

Concentrations above land disposal restrictions: True

Site conditions could threaten: True

Exposed to erosion: True

Site Type: Type 1 on raised ground above floodplain

Figure 4: Cora Output Example (CH2M Hill, 1990)

***** VERSION 3.0 DRAFT *****

DATE: 03/19/98

TIME: 11:16:30

CORA EXPERT SYSTEM

RUN:

RUN BY: Matt Thompson

SITE: Thesis Example

CONTAMINATED AREA: Example

RECOMMENDATIONS FOR HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

GENERAL

- o 504 Site access restrictions
- o 503 Groundwater monitoring
- o 105 Surface water diversion and collection type 1

REMOVAL OPTIONS

- o 201 Soil excavation

LANDFILL for contaminated unsaturated soils

- o Landfill not appropriate for contaminant concentrations above land disposal restrictions

Figure 4(cont.): Cora Output Example (CH2M Hill, 1990)

program.

To better understand the decision tree and questions associated with the containment portion of CORA's structure a flow chart has been provided in Figure 5. The flow chart is setup with the question numbers instead of the actual questions. This was done to help reduce the overall size of the flow chart. The translation between the questions and the question numbers is located in Table 1. Answering Q1, Q2, and Q4 with the answers provided on the flow chart was translation between the questions and the question numbers as given in Table 1. Answering Q1, Q2, and Q4 with the answers provided on the flow chart was required to get within the scope of the project [answers: Q1 → Homogeneous Contaminated Unsaturated soils (HCUS) , Q2 → Containment, Q3 → Volatile Organic Carbon]. Any deviation in this step will place that particular case outside of the scope of the project resulting in a different flow chart not provided or discussed. The remainder of the questions are true false. Appropriate answers will take the user through different paths through the flow chart.

The input data from CORA and the neural network were collected in 10 arbitrary training cases at a time with 5 test cases. These 15 cases are simply the questions and answers associated with data runs within CORA. After the first set of data were collected from CORA, the neural network was established. Table 2 shows the number of cases and how each case was used in each run. This table is used to represent the number of training cases and test cases that were used for every run during the research project. A case was defined as an example of the output from CORA used to either train or test the neural network. A run represented a training session of the neural network that attempts to

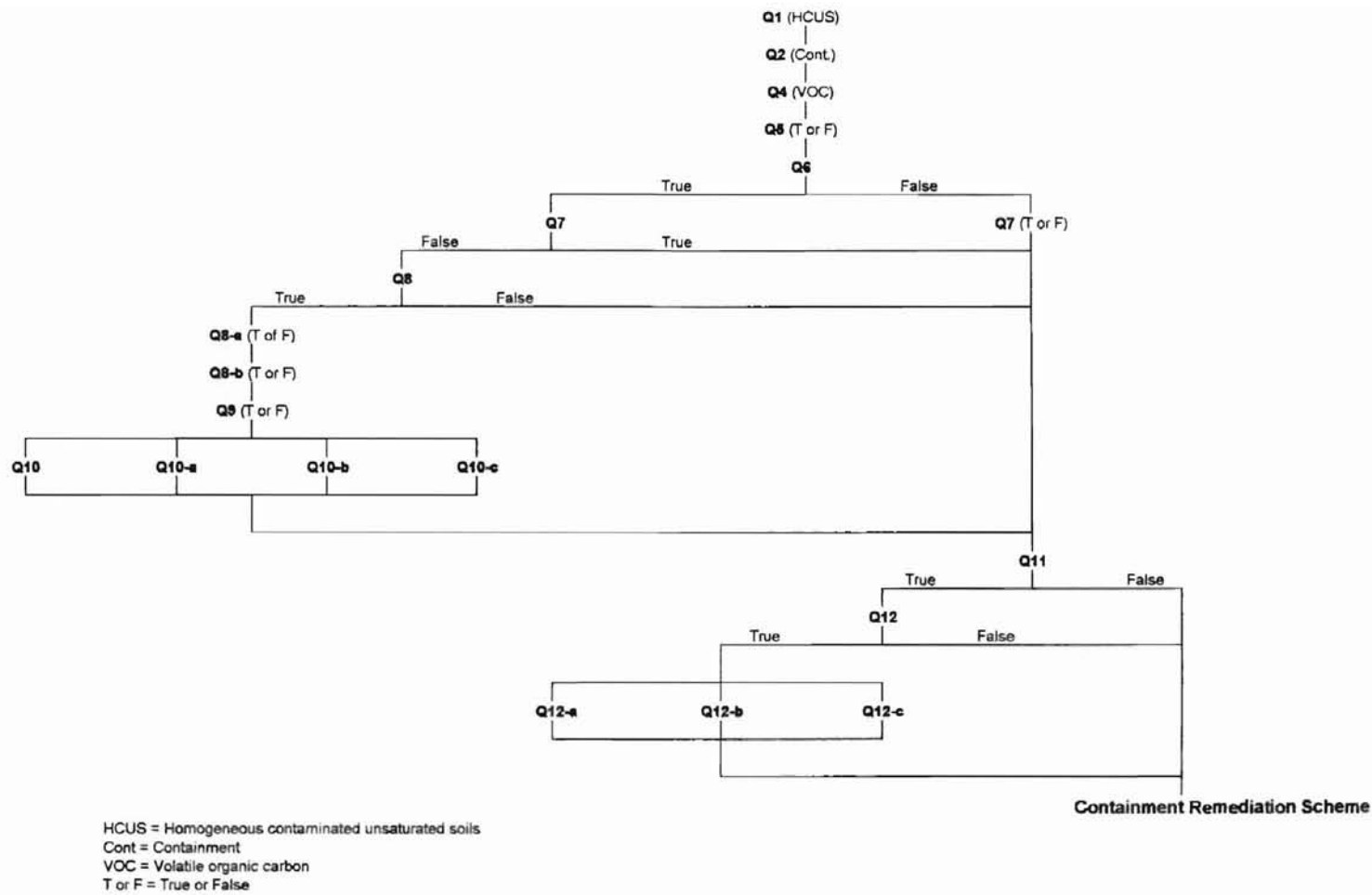


Figure 5: Containment Flow Chart

Table 2: Training and Test Cases from the Neural Network

Run #'s	New Training Cases	Total Training Cases	Test Cases
1,2,3,4,5,6,7, & 8	10	10	5
9,10,11, & 12	9	24	5
13,14,15, & 16	10	39	5
17	10	64	5
18	10	79	5
19	10	94	5
20	10	109	5
21	10	124	5
22	10	139	15

simulation the CORA cases. CORA assigns remediation schemes after every run. This information was used later as target information for the network. Once the training data were determined, a separate group of data were derived to test the neural network once it is trained. This was done by going into CORA and retrieving more data describing different remediation scenarios. The key was that these test data had not previously been placed into the neural network as a training aid. If the test case had been used to train, the neural network will cheat because it already knows the answer from training. Since the neural network would now know the answer from the training, it will use that information for the test case answers and not use the information derived learned through the neural network training structure.

An initial estimate of the size of the spreadsheet was now determined. The overall size of the spreadsheet will be discussed later in section 5.1. The questions that pertained from CORA were placed in the columns. The spreadsheet was setup so that the questions asked by CORA were in individual columns at the top of the spreadsheet. Each data run and its results were recorded in rows with answers to each question in the corresponding column. The answers to the questions from CORA were reported to the neural network

as true or false. A true is represented by the number one in the appropriate square while a false is represented by a zero. A square that does not have any numbers was a question that was not required in that data run and was not used during the training of the network for that particular case. The numbers one and zero in the target and output columns represent whether a particular remediation scheme was acceptable or not. In this case, a one meant that the remediation scheme was appropriate with the given circumstances. A zero meant the opposite: the remediation scheme was not suitable for this application. A total of 154 computer simulations were completed for the containment and remediation alternatives. An example of these with the first 25 simulations from the containment portion, is included in Figure 6. Figure 6 is located inside the back cover in a map folder. The entire spreadsheet containing all 154 simulations are presented in Figure 10 also located in inside the back cover in the map folder. Figure 6 presents the run number as well as a listing of each of the questions previously presented in Table 1 or in the flow chart in Figure 5. Also included in this figure are the twenty targets and outputs associated with each simulation run. They are labeled with the CORA remediation scheme numbers. A complete list of these CORA remediation scheme numbers are located in Table 4.

The columns of the spreadsheet with adjoining questions represent questions from CORA that are directly related. These questions are entered into the neural network structure at the same time as inputs to instruct the neural network of their relationship. The questions that are separated are independent from one another and do not need to be entered into the neural network at the same time. This allows the neural network to know that the questions with adjoining columns are related and should be considered

together during training. After the last CORA question, a section for the targets and outputs was required. One target column was needed for every possible remediation scheme that CORA might suggest for remediating VOC's. The next section to the right in Figure 6 are the output columns. These columns are used by the neural network to produce results based on the trained information. The number of output columns required was equal to the number of target columns. These columns must be labeled in the same order to know the exact answers that the network produced.

4.2 Inputting the Spreadsheet into the Neural Network

Now that the questions and results were entered and the first data are entered the spreadsheet part of the assignment was finished. The spreadsheet must then be dissolved into components. This allows the neural network to understand the information in the spreadsheet. The neural network looked for input columns, target columns, output columns, and a mode column (signals the program that the information for that row was to be used to "Train" or "Test" the network). Training the network told the computer to use a specific row of data for creating pattern associations to predict unknown data runs. The prediction of unknown data runs from the trained network would be classified as a test run.

The first step required to run the neural network program utilized in this effort was to select "initiate working area" from the neural network program menu. Initiate working area officially loads the neural network structure and information onto the spreadsheet. It also allows the user to evaluate the neural network's progress towards

training. The rows that contain data must be highlighted in the spreadsheet and entered into the network with the selection of “Select Rows” from the “Config” menu. This defines the rows of the spreadsheet to be evaluated by the neural network. The input columns were highlighted from the spreadsheet. Input columns are those that were used to represent the answers to CORA’s questions. If the questions were directly related and need to be considered together they must be input simultaneously. As a result, each input column(s) was entered into the network individually by selecting “Add Input Columns” from the “Config” menu. The number of input columns that can be used in this program is practically limitless. Figure 7 shows an example of the screen that was used to load the spreadsheet into the network.

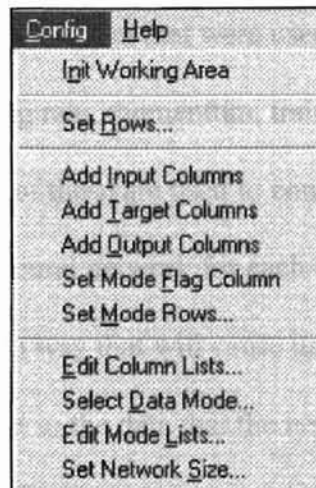


Figure 7-Screen Used for Loading the Spreadsheet

Upon completion of the above steps, all of the information from the CORA questions were loaded into the network. Once all of these input columns were entered, the target data were then entered. As described earlier, the columns that represented the remediation schemes selected by CORA from the answers given are the target columns. All targets from the different scenarios were entered into the network at one time by

means of “Add Input Columns” under the “Config” menu. The Output columns were now highlighted and entered into the neural network through the “Add Output Columns” under the “Config” menu following these activities. The spreadsheet was compatible with the neural network.

4.3 Neural Network Parameters

The parameters of the neural network were the next to be set. These parameters were used to help force the learning curve of the neural network to be smooth and short. In other words, these functions can help remove any points during the training procedure that would cause the neural network to take longer to train. It also allows the acceleration of the training procedure. Those parameters that were used to ease the training procedure include size of the network, learning rate, momentum, training tolerance, and testing tolerance are determined. All five of these are used to control the way in which error is used to correct the weights in the neural network for each training case. The problem is that these parameters can be set in a way that will cause unpredictability. This arises from setting the parameters so close to the data that the network places too much emphasis on matching the data that it forgets to teach itself the pattern association inherent within the data.

The number of layers within the neural network is the first determination to be made. The different neural network evaluated were 2, 3, 4, and 5 layer systems. Each of the systems were tried until they were eliminated for various reasons. A complete list of all of the layer systems that were used during this project is located in the result section

in Table 5. The decisions that were made to determine exactly which layer system was ultimately employed was determined by the amount and type of data. The types of reasons that were used to eliminate systems were that a system could not be trained in a reasonable amount of time and eliminating a system because it did not produce the desired results. The number of neurons remained constant for each layer. They were as follows: input layer → 32 neurons (always the first layer), output layer → 21 neurons (always the last layer), hidden layer 1 → 30 neurons, hidden layer 2 → 45 neurons, hidden layer 3 → 30 neurons. These were chosen by recommendations given by the Neuralyst user guide (Shih, 1994). The number of layers and number of neurons were determined by an iterative process. The figure presented below represents the screen in Neuralyst that was used to set the rest of the network parameters. This screen shows the default values that used by the neural network.

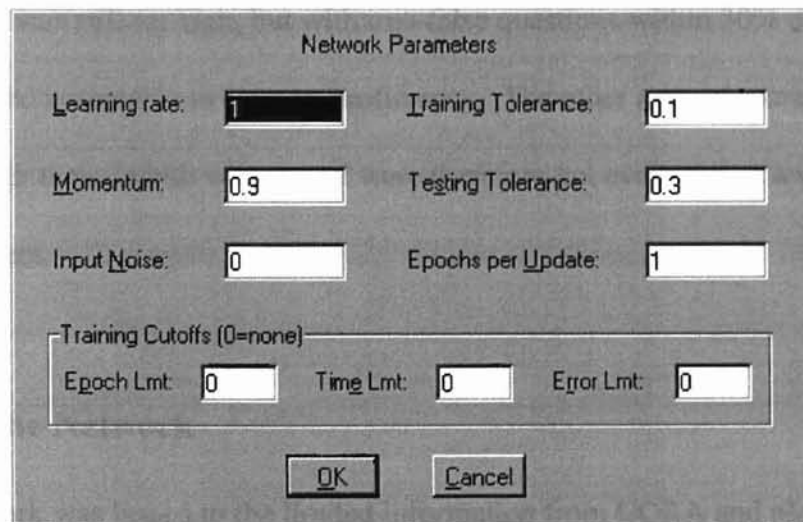


Figure 8 - Neural Network Parameters

Three different scenarios involving the network parameters were evaluated during the containment and treatment training and testing periods. The containment portion of the project looked at only two of these scenarios. The first was accepting the default values

that are listed in Figure 8. The second case accepted the default values above except for changing the training tolerance and testing tolerance. The first case gave results that were not acceptable because the desired results were not achieved. The modified version was used to force a tighter control of the training and testing tolerances. In both cases, the default values were halved. This was an attempt to force both the training and testing of the neural network to work harder to approach the target values. The training tolerance was cut from 0.1 to 0.05. The testing tolerance was lowered from 0.3 to 0.15. The default case was eliminated because it did not produce the 100% recognition rate. The treatment portion also used two cases. They were the final version from the containment portion (modified version → training tolerance 0.05, and testing tolerance → 0.15) plus one that only modified the training tolerance. This third case was developed to reduce the amount of time and control that the modified case exerted on the neural network. The training tolerance was still set high, but with true-false questions within 30% of being accurate considered acceptable to save on proficiency. The other network parameters were accepted with their default values, and were therefore not evaluated as a variable in this research project.

4.4 Training the Network

The network was linked to the limited information from CORA and placed into the initial spreadsheet. The neural network parameters were also initially set to minimize any problems that might arise that would slow down the neural network training procedure. The next step was to train the network with the first ten training and five test

cases that were picked from CORA. The network was then trained with this limited information. The output from the neural network is reported every epoch. An epoch is simply “x” number of training runs that can be defined by the user. It allows the user to check on the progress of the network as it trains. The amount of time that the training run takes is dependent on all the network parameters, the complexity of the network, and the volume of data in the network. During the training procedure the network may never stop training. If this occurs, the training must be stopped and the input data reevaluated. The network parameters must be changed and/or the number of layers or neurons in the network altered. After the above items have evaluated and changed the network can be retrained.

Once the training was finished (the computer screen shows that there is a 100% recognition rate), the network must be used to find the answers to the five test cases defined by using the run/predict portion of the program. The neural network then displays the results of these predictions. For a first run with only a small amount of data, the accuracy will probably not achieve a 100% precision rate. The 100% success rate can be achieved by turning the test cases into training cases and going back into CORA and obtaining additional information for other scenarios. A new column is added to the spreadsheet to accommodate new information that might arise with different scenarios. This research obtained 10 more training cases and 5 more test cases from CORA and further retrained the network. This process was repeated until the success rate of 100% was achieved by the neural network for containment type of remediation schemes. Once 100% was achieved, 10 more test cases were derived from CORA to further the confidence of the neural network.

4.5 Spreadsheet Modification for Treatment Options

The treatment portion of CORA's remediation of VOC's was then explored. Information from CORA was obtained for the various treatment options. This involved the exact same steps as with the containment scenarios. The major difference from the containment and treatment portion were the types of questions asked by CORA and the remediation schemes subsequently suggested. To compensate for the new questions, the existing spreadsheet was extended. A list of the questions that arose from the treatment section are located in Table 3.

Table 3: CORA Treatment Questions*

- What waste types apply to the site (Q1)
- What response action do you wish to consider (Q2)
- What is the hydraulic conductivity of the soil (Q3)
- What types of contaminants in the soil (Q4)
- Will excavation of the contaminants cause environmental or public impacts (Q5)
- Is onsite incineration precluded based on space or local considerations (Q5-a)
- Type of discharge option either water reinjection (Q10)-water infiltration (Q10-a)-discharge to POTW (Q-10b)-discharge to surface water (Q10-c)
- What is the hydraulic conductivity of saturated zone (Q10-d)
- Is the water table greater than 5 feet below surface (Q10-e)
- Is the ash a hazardous waste (Q10-f)
- Is conc. of containment in ash above land disposal (Q10-g)
- Is an onsite RCRA landfill for solidified ash reasonable (Q10-h)
- What types of contamination in the leachate from landfill (Q10-i)
- Would a shallow aquifer preclude a below grade landfill for solidified ash (Q10j)
- Is the solidified material landfill footprint in a 100-year floodplain (Q10-k)
- Could site conditions threaten health or safety of unauthorized visitors (Q11)
- Are exposed soils on the site exposed to erosion (Q12)
- Pick the location of the site above floodplain (Q12-a)-at base of hill above floodplain (Q12-b)-in floodplain (Q12-c)

* The questions are not sequentially numbered due to other questions that resulted in the containment section of CORA

A complete flow chart of the treatment section of CORA is presented in Figure 9. Figure 9 is much like Figure 5, the number of the appropriate questions rather than the actual

questions are included in this flow chart. The actual question numbers can be replaced with the questions with the help of Table 3. The first three questions must be answered appropriately to fall under the treatment scope of this project. These answers to these questions are homogeneous contaminated unsaturated soils (Q1), treatment (Q2), and volatile organic carbons (Q3). The remainder of the questions are true false. Depending on answers given by user the expert system will determine the final path that is used through the flow chart for a selected scenario. The different remediation schemes associated with treatment were added to the target and to the output columns. The increased information caused the neural network parameters to be reevaluated and adjusted. The spreadsheet was cleared and reloaded into the neural network framework. After this was accomplished, the network could be trained and tested as with the containment portion of the network. This process for treatment was repeated until the 100% success rate was accomplished.

4.6 Updating Cost of CORA's Remediation Technologies

The neural network was now complete for containment and treatment options and was capable of producing CORA compatible answers for either containment or treatment of VOC's. The costs portion of the expert system required attention. The goal of this portion of the project was to give updated costs from CORA 1990 to include a more current data set. Several ways were looked at for this step. The first was to try to use data from CORA and simply apply a price of inflation to adjust the cost to 1998

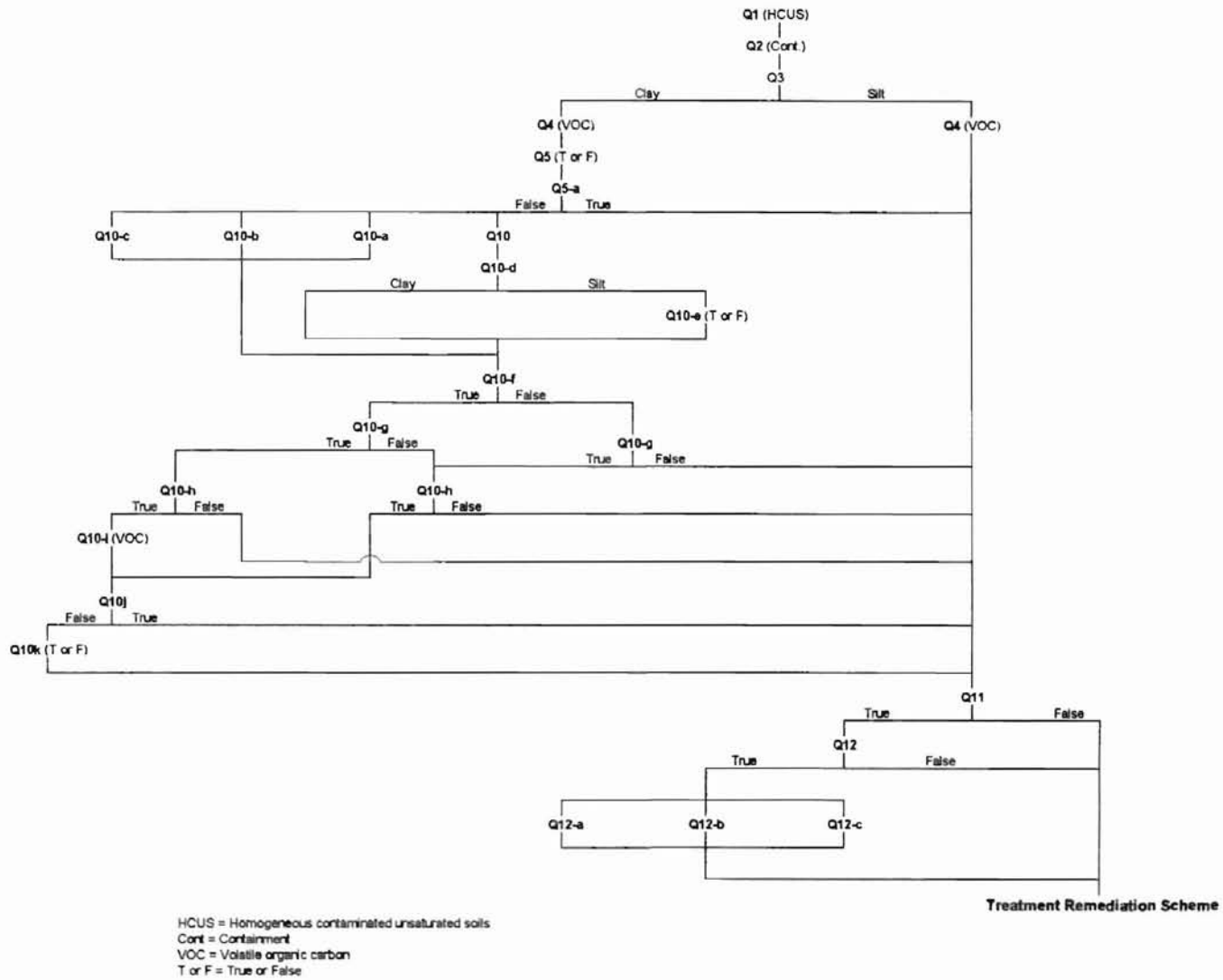


Figure 9: Treatment Flow Chart

dollars. The problem with this was that CORA does not give unit costs for the remediation schemes. This meant that volumes of contamination had to be known to give a price for that scheme. An exact amount of contamination could not be determined for a base case since every remediation scheme in CORA requires a different minimum amount of contamination. Defining the current prices based on a unit cost was eventually selected as the method of choice. This allowed every site to be economically evaluated by eventually knowing the unit cost from the neural network and the volume of contaminated material from the engineers studies. The costs of the remediation schemes came from several sources. They include the Environmental Protection Agency (1998), Ground-Water Remediation and Analysis Center (1998), environmental design engineers, city officials, and landfill designers.

5.0 Results from the Neural Network

The results of this research project come in two forms. Those involved the actual setup of the input spreadsheet containing both the questions and resultant answers from CORA and the actual training of the network required to meet the 100% recognition rate as defined by the scope of the research project. In both cases, the process that was used to arrive at the correct answer was a result of adjustments and iterations.

5.1 Model Dimensions

The actual spreadsheet turned out to be quite large and extensive. The total size of the spreadsheet excluding the neural network weights was 88 columns by 156 rows. The complete spreadsheet is located in Figure 10 inside the back cover in map folder 1. The total number of trial cases that it took to train and test the neural network for both containment and treatment from CORA's information was 154. Fifty-four of those runs were containment, while the remaining 100 were used to train the network for the treatment section of the project. The containment portion of CORA required up to 19 questions while the treatment portion of the project required answers up to 23 questions. The exact number of questions that were answered by the user is highly site dependent. A complete list of the questions from CORA and the neural network infrastructure have previously been presented and discussed in Chapter 4. The flow sheet that incorporates both treatment and containment in the decision tree of CORA is located in Figure 11. The total numbers of different types of remediation schemes developed by CORA for the

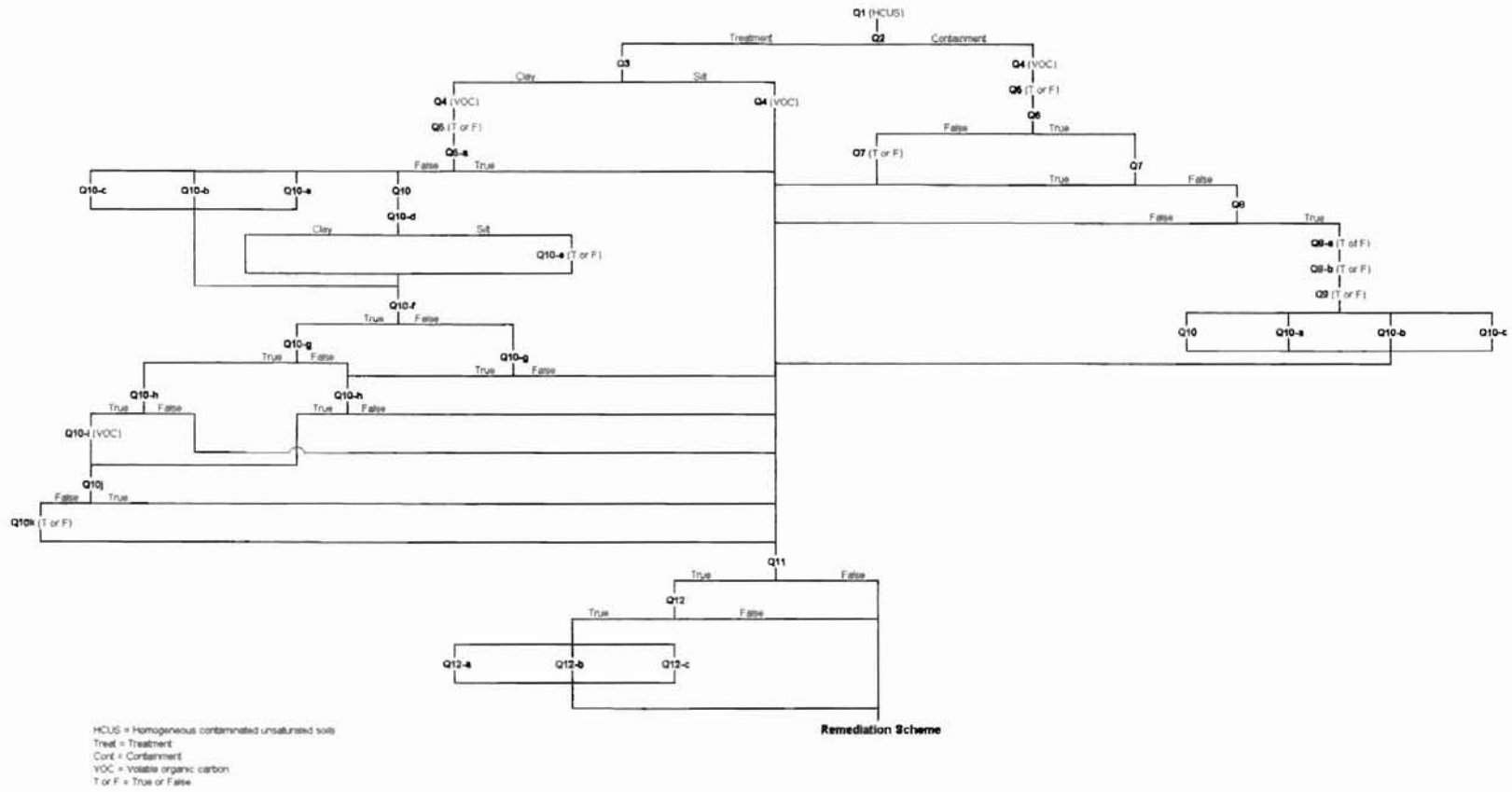


Figure 11: Treatment and Containment Flow Chart (Combined)

this project was 20. Containment schemes used 11 of the 20 schemes. The treatment portion of the project required all 20 of the different types of schemes that were developed by CORA. The table (Table 4) on the following page represents the remediation schemes used by CORA. It also shows which schemes were used by each remediation type. Several of the different schemes were found in both the containment and treatment options.

5.2 Training the Network

The number of different runs that were required by the neural network to simulate the CORA's responses following training was 22. Recall, that this effort was not designed to determine the minimum amount of data required to achieve the 100% recognition rate. Rather, the project goals were to achieve the 100% success rate of both the containment and treatment portion of CORA with VOC's as the chemical of concern. Containment options required 16 of the training runs, while the treatment portion only required six runs. The containment scenarios needed more runs because they were used to develop the number of neural network layers and the network parameters which were then utilized by the treatment simulations. The number of runs and data samples were determined by trial and error. If the first trial of the data achieved the desired goal, then the number of data samples would be sufficient for that particular case. In both cases 10 additional cases were evaluated to further the precision of the network.

The framework for the 22 runs was as follows. Sixteen runs were used for the containment portion while 6 were used for the treatment portion. Table 5 displays the run

number, number of samples, number of layers, neurons, parameters, and results for each of the 22 runs.

Table 4: Remediation Schemes used in Containment and Treatment

REMEDIATION SCHEME CORA #	NAME	CONTAINMENT	TREATMENT
105	Surface Water Diversion/Collection	X	X
201	Soil Excavation	X	X
301	Onsite Incineration		X
302	Offsite Incineration		X
305	Soil Vapor Extraction		X
306	Flaring		X
307	Air Stripping		X
308	Vapor Phase Carbon		X
312	Ion Exchange		X
316	Solidification		X
317	In-situ Stabilization		X
401	Offsite RCRA Landfill	X	X
402	Onsite RCRA Landfill - Above Grade	X	X
403	Offsite RCRA Landfill - Below Grade	X	X
404	Offsite Solid Waste Landfill	X	X
405	Discharge to POTW	X	X
406	Discharge to Surface Water	X	X
407	Water ReInjection		X
503	Groundwater Monitoring	X	X
504	Site Access Restrictions	X	X

"X" denotes when a particular remediation scheme could have been employed by CORA

The network became so large that the three and four layer systems could not

Table 5: Results of the Training Runs

Run Number	Number of Samples	Number of Layers	Neurons per Layer	*Network Parameters	Results (correct)	Comments
1	10 train 5 test	2	32,21	Default Values	93%	Need more data
2	10 train 5 test	2	32,21	Training Tol. = 0.05 Testing Tol. = 0.15	88%	Need more data
3	10 train 5 test	3	32,30,21	Default Values	94%	Need more data
4	10 train 5 test	3	32,30,21	Training Tol. = 0.05 Testing Tol. = 0.15	90%	Need more data
5	10 train 5 test	4	32,30,45,21	Default Values	93%	Need more data
6	10 train 5 test	4	32,30,45,21	Training Tol. = 0.05 Testing Tol. = 0.15	89%	Need more data
7	10 train 5 test	5	32,30,45,30,21	Default Values	91%	Need more data
8	10 train 5 test	5	32,30,45,30,21	Training Tol. = 0.05 Testing Tol. = 0.15	90%	Need more data
9	24 train 5 test	2	32,21	Training Tol. = 0.05 Testing Tol. = 0.15	96%	Need more data
10	24 train 5 test	3	32,30,21	Training Tol. = 0.05 Testing Tol. = 0.15	93%	Need more data
11	24 train 5 test	4	32,30,45,21	Training Tol. = 0.05 Testing Tol. = 0.15	92%	Need more data
12	24tr 5 test	5	32,30,45,30,21	Training Tol. = 0.05 Testing Tol. = 0.15	95%	Need more data
13	39 train 5 test	2	32,21	Training Tol. = 0.05 Testing Tol. = 0.15	---	Would not train eliminated 2 layer networks
14	39 train 5 test	3	32,30,21	Training Tol. = 0.05 Testing Tol. = 0.15	96%	Need more data
15	39 train 5 test	4	32,30,45,21	Training Tol. = 0.05 Testing Tol. = 0.15	99%	Need more data
16	39 train 5 test	5	32,30,45,30,21	Training Tol. = 0.05 Testing Tol. = 0.15	100%	100% training for containmnet --10 more cases to test
17	64 train 5 test	5	32,30,45,30,21	Training Tol. = 0.05	59%	Need more data
18	79 train 5 test	5	32,30,45,30,21	Training Tol. = 0.05	90%	Need more data
19	94 train 5 test	5	32,30,45,30,21	Training Tol. = 0.05	94%	Need more data
20	109 train 5 test	5	32,30,45,30,21	Training Tol. = 0.05	90%	Need more data
21	124 train 5 test	5	32,30,45,30,21	Training Tol. = 0.05	96%	Need more data
22	139 train 5 test	5	32,30,45,30,21	Training Tol. = 0.05	100%	100% training for treatment --10 more cases to test

* Default network parameters are as follows: Learning Rate = 1, Momentum = 0.9, Input Noise = 0, Testing Tolerance = 0.3, Training Tolerance = 0.1, and 1 Epochs per update → These values are used unless a modified version was looked at in this column. In this case, the default values were used except for the changes noted in this column.

handle the number of variables and information in a routine manner. The run time for three training sessions was stopped at two hours without 100% training rate. This was deemed an unacceptable time limit. Without the training achieving the 100% recognition rate, the prediction of the test cases is unpredictable. Therefore, the 2, 3, and 4 layer systems were eliminated after the completion of containment section. The desired goal of 100% was achieved with the five layer system in the containment section so it was used for the training of the treatment section. In all cases, if a smaller or larger layer system is needed (or tried) the neural network can be altered and the network reloaded instantly.

The final network parameters were all default values except the training tolerance. As the chart shows, the training tolerance was set to 0.05. This forced the neural network to be within 0.05 of the actual number during the training procedure. This tighter control during training helped force the training numbers closer to the target values. The other network variables were not modified because upon further review the default values were in the appropriate range to achieve the best results in a short amount of time.

5.3 Cost Results

The cost that were associated with the remediation schemes from CORA in this project were updated with basic unit cost. The table on the following page (Table 6) shows the remediation scheme name and CORA number with the updated cost. The basic unit costs are in the form of price per ton, price per cubic yard, price per pound, price per gallon, and price per size of machine. The exact conditions of the site will ultimately determine the cost of the project.

Table 6: Updated Costs for CORA

REMEDIATION SCHEME CORA #	NAME	UNIT COSTS
105	Surface Water Diversion/Collection	\$10,963/Ac *
201	Soil Excavation	Approximately \$2-\$5/yd ³
301	Onsite Incineration	\$164 -- \$730/Tons
302	Offsite Incineration	\$200 -- \$1,000/Tons
305	Soil Vapor Extraction	\$100/Tons
306	Flaring	\$300/hole for pipes
307	Air Stripping	Depends on Electricity costs -- Requires 1.5 hp/Foot of Stripping
308	Vapor phase Carbon	\$1,000 -- \$40,000 for the Machine Carbon = \$2 -\$3/lb
312	Ion Exchange	\$0.30 -- \$0.80/1000 gal Treated
316	Solidification	\$100/Ton Including Excavation
317	In-situ Stabilization	Shallow -- \$40-\$60/yd ³ Deep -- \$150-\$250/yd ³
401	Offsite RCRA Landfill	\$15/Yard - Excludes Transportation
402	Onsite RCRA Landfill - Above Grade	\$500-\$1140/cy Range from 7000- 220000cy *
403	Offsite RCRA Landfill - Below Grade	\$490-\$1121/cy Range from 7000- 220000cy *
404	Offsite Solid Waste Landfill	\$4.00/cy plus \$1.50/ton
405	Discharge to POTW	\$5.25/gal First 1000 gal, \$2.00/gal After
406	Discharge to Surface Water	NPDES Permit = \$7,000 **
407	Water Reinjection	\$1.00/gal Haz, \$0.55/gal Non Haz Excluding Pump Truck
503	Groundwater Monitoring	\$2,000/well/month plus Quarterly Monitoring
504	Site Access Restrictions	\$28.50/ft Includes Fencing and Signs*

* These costs came from CORA and were updated from 1990 dollars to 1998 dollars with a factor of inflation of 3%, ** This price depends on the city that issues the NPDES permit (this price is for Sand Springs, OK).

5.4 Examples to Compare the Neural Network to CORA

The following example presents a comparison of the output from the neural network model with those generated by CORA. This hypothetical site is typical of those found in Oklahoma. The example was broken down into five cases. Each case was similar, but with minor changes to show the process of working through the decision tree (the decision tree is located in the flow chart in Figure 11) and arriving at different remediation schemes. Two cases (Case 1 and 2) involved a containment scenario for remediation, while the remaining three cases (Cases 3, 4, and 5) were treatment scenarios. Case 2 was very similar to that used as an example of the CORA output in section 4.1 in Figure 4.

The following information was used as background information for the site and was used in all five of the test cases:

- A homogeneous contaminated unsaturated soil
- Non-hazardous volatile organic carbons (VOC's) as the chemical of concern
- Clayey soil
- A site where the physical nature was dangerous to trespassers
- Exposed soils that may erode
- A site that can be excavated
- A small site that restricts any type of onsite landfill
- A site that is located above the flood plain
- A site that is located near surface water.

The information that was variable between all the cases is listed below.

- The type of remediation scheme employed (containment/treatment)
- Concentration of the soil (above/below land disposal restrictions)
- Incineration (yes/no)
- Concentration of ash (above/below land disposal restrictions)

Case 1 looked at a containment scheme where the concentrations of the soil were not above land disposal restrictions. Case 2 investigated the same scenario except that

the soils were above the land disposal restrictions. The first treatment scheme, case 3, did not use a incineration. While case 4 looked was a treatment scheme that employed incineration. The ash in this case was not above the landfill disposal restrictions. The last case (case 5) is the same as case 4 except the ash was above the land disposal restrictions.

Questions asked by CORA that are used for inputs to the neural network are as follows:

CASE 1

Q1: What waste types apply to the site? *Homogeneous contaminated unsaturated soils*

Q2: What response action do you wish to consider? *Containment*

Q4: What types of contaminants are in the soil? *VOC's*

Q5: Will excavation of the contaminants not cause environmental or public impacts? *True*

Q6: Is the contaminated soil a hazardous substances? *False*

Q7: Is the contaminated soil concentration above land disposal restrictions? *False*

Q11: Could site conditions threaten health or safety of unauthorized visitors? *True*

Q12: Are exposed soils on the site exposed to erosion? *True*

Q12-a.b.c: Pick the location of the site: *Above Floodplain (Q12-a)*

These questions can be followed through the flow chart in Figure 11. The actual questions instead of question numbers are located Table 1.

The results of these inputs for CORA and the neural network were the same. They included remediation schemes 504-site restrictions, 503-groundwater monitoring, 105-surface water diversion, 201-soil excavation, and 404-offsite solid waste landfill. The number preceding the activity refers to the CORA module describing that activity. The CORA results are located on the following two pages in Figure 12. The neural network results are located in Figure 17.

***** VERSION 3.0 DRAFT *****

DATE: 03/25/98
TIME: 20:15:36

CORA EXPERT SYSTEM

RUN: Case 1
RUN BY: Matt Thompson
SITE: Oklahoma Site, Case 1
CONTAMINATED AREA: Case 1

WASTE TYPE: HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

INPUT

Response type: Containment
Soil contaminant: Volatile organic compounds
Excavation acceptable: True
Material in question is hazardous: False
Concentrations above land disposal restrictions: False
Site conditions could threaten: True
Exposed to erosion: True
Site Type: Type 1 on raised ground above floodplain

Figure 12: Example Problem - Case 1 CORA Results (CH2M Hill, 1990)

***** VERSION 3.0 DRAFT *****

DATE: 03/25/98

TIME: 20:15:36

CORA EXPERT SYSTEM

RUN: Case 1
RUN BY: Matt Thompson
SITE: Oklahoma Site, Case 1
CONTAMINATED AREA: Case 1

RECOMMENDATIONS FOR HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

GENERAL

- o 504 Site access restrictions
- o 503 Groundwater monitoring
- o 105 Surface water diversion and collection type 1

REMOVAL OPTIONS

- o 201 Soil excavation

LANDFILL for contaminated unsaturated soils

- o 404 Offsite solid waste landfill

Figure 12(cont.): Example Problem - Case 1 CORA Results(CH2M Hill, 1990)

CASE 2

Q1: What waste types apply to the site? *Homogeneous contaminated unsaturated soils*

Q2: What response action do you wish to consider? *Containment*

Q4: What types of contaminants are in the soil? *VOC's*

Q5: Will excavation of the contaminants not cause environmental or public impacts? *True*

Q6: Is the contaminated soil a hazardous substance? *False*

Q7: Is the contaminated soil concentration above land disposal restrictions? *False*

Q11: Could site conditions threaten health or safety of unauthorized visitors? *True*

Q12: Are exposed soils on the site exposed to erosion? *True*

Q12-a,b,c: Pick the location of the site: *Above Floodplain (Q12-a)*

These questions can be followed through the flow chart in Figure 11. The actual questions instead of question numbers are located Table 1.

The results of this case were 504-site access restrictions, 503-groundwater monitoring, 105-surface water diversion, and 201-soil excavation. The results were also the same between CORA and the neural network. Case 2 CORA results are located on the following two pages in Figure 13. Figure 17 shows the neural network results. The results of this case when compared to case 1 show that offsite solid waste incineration is not needed.

CASE 3

Case 3 was the first of the three cases that looked at treatment versus containment for site remediation. The background information mentioned earlier in this section was the same for this case. This case excludes onsite incineration. Questions that were asked by CORA for this case were:

***** VERSION 3.0 DRAFT *****

DATE: 03/25/98
TIME: 20:18:00

CORA EXPERT SYSTEM

RUN: Case 2
RUN BY: Matt Thompson
SITE: Oklahoma Site, Case 2
CONTAMINATED AREA: Case 2

WASTE TYPE: HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

INPUT

Response type: Containment
Soil contaminant: Volatile organic compounds
Excavation acceptable: True
Material in question is hazardous: False
Concentrations above land disposal restrictions: True
Site conditions could threaten: True
Exposed to erosion: True
Site Type: Type 1 on raised ground above floodplain

Figure 13: Example Problem - Case 2 CORA Results (CH2M Hill, 1990)

***** VERSION 3.0 DRAFT *****

DATE: 03/25/98
TIME: 20:18:00

CORA EXPERT SYSTEM

RUN: Case 2
RUN BY: Matt Thompson
SITE: Oklahoma Site, Case 2
CONTAMINATED AREA: Case 2

RECOMMENDATIONS FOR HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

GENERAL

- o 504 Site access restrictions
- o 503 Groundwater monitoring
- o 105 Surface water diversion and collection type 1

REMOVAL OPTIONS

- o 201 Soil excavation

LANDFILL for contaminated unsaturated soils

- o Landfill not appropriate for contaminant concentrations above
land disposal restrictions

Figure 13(cont.): Example Problem - Case 2 CORA Results (CH2M Hill, 1990)

Q1: What waste types apply to the site? *Homogeneous contaminated unsaturated soils*

Q2: What response action do you wish to consider? *Treatment*

Q3: What is the hydraulic conductivity of the soil? *Clay*

Q4: What types of contaminants are in the soil? *VOC's*

Q5: Will excavation of the contaminants not cause environmental or public impacts? *True*

Q5-a: Is on-site incineration option precluded based on space or local considerations? *True*

Q11: Could site conditions threaten health or safety of unauthorized visitors? *True*

Q12: Are exposed soils on the site exposed to erosion? *True*

Q12-a,b,c: Pick the location of the site: *Above Floodplain (Q12-a)*

These questions can be followed through the flow chart in Figure 11. The actual questions instead of question numbers are located Table 3.

Results from case 3 were 504-site access restrictions, 503-groundwater monitoring, 105 surface water diversion, 201-soil excavation, and 302 offsite incineration. The results from CORA are located in the following figure (Figure 14). The neural network results are posted in Figure 17. These results differed from the containment schemes completed in cases 1 and 2.. The difference in this case and the previous two cases was the offsite incineration feature.

CASE 4

This case was a treatment scheme that allowed onsite incineration. The questions that were asked by CORA were:

Q1: What waste types apply to the site? *Homogeneous contaminated unsaturated soils*

Q2: What response action do you wish to consider? *Treatment*

Q3: What is the hydraulic conductivity of the soil? *Clay*

***** VERSION 3.0 DRAFT *****

DATE: 03/25/98
TIME: 20:20:44

CORA EXPERT SYSTEM

RUN: Case 3
RUN BY: Matt Thompson
SITE: Oklahoma Site, Case 3
CONTAMINATED AREA: Case 3

WASTE TYPE: HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

INPUT

Response type: Treatments
Soil description: Clay
Excavation acceptable: True
Soil contaminant: Volatile organic compounds
Onsite incineration is precluded: True
Site conditions could threaten: True
Exposed to erosion: True
Site Type: Type 1 on raised ground above floodplain

Figure 14: Example Problem - Case 3 CORA Results (CH2M Hill, 1990)

***** VERSION 3.0 DRAFT *****

DATE: 03/25/98
TIME: 20:20:44

CORA EXPERT SYSTEM

RUN: Case 3
RUN BY: Matt Thompson
SITE: Oklahoma Site, Case 3
CONTAMINATED AREA: Case 3

RECOMMENDATIONS FOR HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

GENERAL

- o 504 Site access restrictions
- o 503 Groundwater monitoring
- o 105 Surface water diversion and collection type 1

REMOVAL OPTIONS

- o 201 Soil excavation

INCINERATION

- o 302 Offsite incineration

Figure 14(cont.): Example Problem - Case 3 CORA Results (CH2M Hill, 1990)

Q4: What types of contaminants are in the soil? *VOC's*

Q5: Will excavation of the contaminants not cause environmental or public impacts? *True*

Q5-a: Is on-site incineration option precluded based on space or local considerations? *False*

Q10-a,b,c,d: Type of discharge option? *Discharge to surface water (Q10-d)*

Q10-f: Is the ash a hazardous waste? *False*

Q10-g: Is the concentration in ash above land disposal requirements? *False*

Q11: Could site conditions threaten health or safety of unauthorized visitors? *True*

Q12: Are exposed soils on the site exposed to erosion? *True*

Q12-a,b,c: Pick the location of the site: *Above Floodplain (Q12-a)*

These questions can be followed through the flow chart in Figure 11. The actual questions instead of question numbers are located Table 3.

Results for case 4 were as follows: 504-site access restrictions, 503 groundwater monitoring, 105-surface water diversion, 201-soil excavation, 302-offsite incineration, 301-onsite incineration, 312-ion exchange, 406-discharge to surface water, and 404-offsite solid waste landfill. The actual CORA results are located on the following two pages in Figure 15 while the neural network results are located in Figure 17. More than the others, this case gave the user an assorted list of remediation schemes to consider.

CASE 5

Case 5 is the last example completed. Similar to case 4, it differed in that contamination concentration of the ash exceeded the land disposal restrictions.

The questions that were asked in case 5:

Q1: What waste types apply to the site? *Homogeneous contaminated unsaturated soils*

***** VERSION 3.0 DRAFT *****

DATE: 03/25/98
TIME: 20:22:33

CORA EXPERT SYSTEM

RUN: Case 4
RUN BY: Matt Thompson
SITE: Oklahoma Site, Case 4
CONTAMINATED AREA: Case 4

WASTE TYPE: HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

INPUT

Response type: Treatments
Soil description: Clay
Excavation acceptable: True
Soil contaminant: Volatile organic compounds
Onsite incineration is precluded: False
Discharge options: Discharge to surface water

USER RESPONSES FOR ash

Material in question is hazardous: False
Concentrations above land disposal restrictions: False

USER RESPONSES FOR Homogeneous contaminated unsaturated soils

Site conditions could threaten: True
Exposed to erosion: True
Site Type: Type 1 on raised ground above floodplain

Figure 15: Example Problem - Case 4 CORA Results (CH2M Hill, 1990)

***** VERSION 3.0 DRAFT *****

DATE: 03/25/98
TIME: 20:22:33

CORA EXPERT SYSTEM

RUN: Case 4
RUN BY: Matt Thompson
SITE: Oklahoma Site, Case 4
CONTAMINATED AREA: Case 4

RECOMMENDATIONS FOR HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

GENERAL

- o 504 Site access restrictions
- o 503 Groundwater monitoring
- o 105 Surface water diversion and collection type 1

REMOVAL OPTIONS

- o 201 Soil excavation

INCINERATION

Either

- o 302 Offsite incineration

Or

- o 301 Onsite incineration

o See discharge recommendations for treated scrubber blowdown.

o See landfill recommendations for ash

- o 312 Ion exchange for discharge options other than to POTW

DISCHARGE

- o 406 Discharge to surface water

LANDFILL for ash

Either

- o 404 Offsite solid waste landfill

Or

- o Use ash to backfill excavation

Figure 15(cont.): Example Problem - Case 4 CORA Results (CH2M Hill, 1990)

- Q2: What response action do you wish to consider? *Treatment*
- Q3: What is the hydraulic conductivity of the soil? *Clay*
- Q4: What types of contaminants are in the soil? *VOC's*
- Q5: Will excavation of the contaminants not cause environmental or public impacts? *True*
- Q5-a: Is on-site incineration option precluded based on space or local considerations? *False*
- Q10-a,b,c,d: Type of discharge option? *Discharge to surface water (Q10-d)*
- Q10-f: Is the ash a hazardous waste? *False*
- Q10-g: Is the concentration in ash above land disposal requirements? *True*
- Q10-h: Is an onsite RCRA landfill for solidified ash reasonable? *False*
- Q11: Could site conditions threaten health or safety of unauthorized visitors? *True*
- Q12: Are exposed soils on the site exposed to erosion? *True*
- Q12-a,b,c: Pick the location of the site: *Above Floodplain (Q12-a)*

These questions can be followed through the flow chart in Figure 11. The actual questions instead of question numbers are located Table 3.

Case 5 involved more remediation alternatives: 504-site access restrictions, 503 groundwater monitoring, 105-surface water diversion, 201-soil excavation, 302-offsite incineration, 301-onsite incineration, 312-ion exchange, 406-discharge to surface water, 317-in-situ stabilization, 316-solidification, and 401-offsite RCRA landfill. CORA outputs are located in Figure 16 while the neural network results are located in Figure 17. These results differ from the other four cases because of the change in the inputs.

Table 7 summarizes the results of all five cases for both CORA and the neural network inputs. The results were directly compared in every case. The results indicate a 100% matching of remediation schemes between these two programs. This indicated that the neural network was now completely trained and ready for use on projects that fall under the same conditions that exist in this research project.

***** VERSION 3.0 DRAFT *****

DATE: 03/25/98
TIME: 20:25:38

CORA EXPERT SYSTEM

RUN: Case 5
RUN BY: Matt Thompson
SITE: Oklahoma Site, Case 5
CONTAMINATED AREA: Case 5

WASTE TYPE: HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

INPUT

Response type: Treatments
Soil description: Clay
Excavation acceptable: True
Soil contaminant: Volatile organic compounds
Onsite incineration is precluded: False
Discharge options: Discharge to surface water

USER RESPONSES FOR ash

Material in question is hazardous: False
Concentrations above land disposal restrictions: True
Onsite landfill reasonable: False

USER RESPONSES FOR Homogeneous contaminated unsaturated soils

Site conditions could threaten: True
Exposed to erosion: True
Site Type: Type 1 on raised ground above floodplain

Figure 16: Example Problem - Case 5 CORA Results (CH2M Hill, 1990)

***** VERSION 3.0 DRAFT *****

DATE: 03/25/98
TIME: 20:25:38

CORA EXPERT SYSTEM

RUN: Case 5
RUN BY: Matt Thompson
SITE: Oklahoma Site, Case 5
CONTAMINATED AREA: Case 5

RECOMMENDATIONS FOR HOMOGENEOUS CONTAMINATED UNSATURATED SOILS

GENERAL

- o 504 Site access restrictions
- o 503 Groundwater monitoring
- o 105 Surface water diversion and collection type 1

REMOVAL OPTIONS

- o 201 Soil excavation

INCINERATION

Either

- o 302 Offsite incineration

Or

- o 301 Onsite incineration
- o See discharge recommendations for treated scrubber blowdown.
- o See landfill recommendations for ash
- o 312 Ion exchange for discharge options other than to POTW

DISCHARGE

- o 406 Discharge to surface water

LANDFILL for ash

Either

- o 317 In-situ stabilization

Or

- o 316 Solidification
- o 401 Offsite RCRA landfill

Figure 16(cont.): Example Problem - Case 5 CORA Results (CH2M Hill, 1990)

RUN	Q1	Q2	Q3	Q4	Q5	Q5-a	Q6	Q7	Q10-c	Q10-f	Q10-g	Q10-h	Q11	Q12	Q12-a
Case 1	1	1		1	0		0	0					1	1	1
Case 2	1	0		1	0		0	1					1	1	1
Case 3	1	0		1	0		0	0					1	1	1
Case 4	1	0		1	0		0	0					1	1	1
Case 5	1	0		1	0		0	0					1	1	1

Figure 17: Neural Network results for the Example Problem

Q1 HCUS Select all waste types that apply to the site in question

Q2 CONT TREAT Select a response action that you wish to consider for the appropriate waste type

Q3 CLAY SILT Hydraulic Conductivity - Select what best describes the soils

Q4 VOC Select all types of contaminants in the soil

Q5 Excavation of contaminated soils will result in acceptable environmental or public impacts

Q5-a Onsite incineration is precluded based on space (2-1/2 acre minimum clean space) and/or locational considerations

Q6 Contaminated unsaturated soil in question for landfills is a hazardous waste

Q7 Contaminated unsaturated soil concentration is likely to be above the land disposal restrictions

Q10-c Discharge Option - discharge to surface water

Q10-f Ash is a hazardous waste

Q10-g Concentration of contaminant in ash is above land disposal restrictions

Q10-h Onsite RCRA landfill for solidified ash is reasonable based on available clean space, landuse, and locational

Q11 Site conditions could threaten the health and/or safety of unauthorized site visitors

Q12 Exposed soils on the site that are exposed to erosion

Q12-a Type 1 - on raised ground above floodplain

Run	Targets											
	504	503	105	201	301	302	312	316	317	401	404	406
Case 1	1	1	1	1	0	0	0	0	0	0	0	0
Case 2	0	1	1	1	0	0	0	0	0	0	0	0
Case 3	0	1	1	0	0	0	0	0	0	0	0	0
Case 4	1	1	1	1	0	0	0	0	0	1	0	0
Case 5	0	1	0	0	0	0	0	0	0	0	1	0

Run	Outputs											
	504	503	105	201	301	302	312	316	317	401	404	406
Case 1	1	1	1	1	0	0	0	0	0	0	0	0
Case 2	0	1	1	1	0	0	0	0	0	0	0	0
Case 3	0	1	1	0	0	0	0	0	0	0	0	0
Case 4	1	1	1	1	0	0	0	0	0	1	0	0
Case 5	0	1	0	0	0	0	0	0	0	0	1	0

Figure 17 (cont.): Neural Network results for the Example Problem

Table 7: Comparison of Results for the Example

Case Number	Results	Remediation Number											
		105	201	301	302	312	316	317	401	404	406	503	504
1	CORA	x	x							x		x	x
1	NN	x	x							x		x	x
2	CORA	x	x									x	x
2	NN	x	x									x	x
3	CORA	x	x		x							x	x
3	NN	x	x		x							x	x
4	CORA	x	x	x	x	x				x	x	x	x
4	NN	x	x	x	x	x				x	x	x	x
5	CORA	x	x	x	x	x	x	x	x		x	x	x
5	NN	x	x	x	x	x	x	x	x		x	x	x

NN = Neural Network, The actual remediation names instead of the numbers are located in Table 4. The actual results for this chart are located in Figure 12, 13, 14, 15, 16, and 17.

6.0 SUMMARY

The scope of this project was to accurately predict examples (100% recognition rate) of remediation types (based on CORA) in a timely manner for any site in Oklahoma with the criteria:

- it is a homogeneous contaminated saturated soil,
- VOC's are the only contaminant, and
- the soil type is either clay or silt.

These goals were accomplished by taking the slightly dated information from CORA and modeling it with the use of a neural network. The CORA information that was used to establish the neural network model had to be set up within the structure of a spreadsheet. Once the spreadsheet was developed, the data was loaded into the neural network where it could be run for the training and predicting stages of the project.

This neural network ultimately required 54 data cases and 16 training runs to accurately produce the 100% pattern recognition rate for the containment section of this research project. Some of the 16 runs were used to determine the size of the neural network and the number of neurons. The ultimate size of the neural network was dependent on both the treatment and containment portion of the project. The treatment portion of the project involved more information from CORA that required almost twice as many data samples as the containment section (100 data cases) to be completely trained. The number of runs required for the treatment section was six.

The final size of the neural network structure was five layers with the following numbers of neurons per layer 32, 30, 45, 30, and 21 respectively for a total of 158 neurons. The final values that were used for the network parameters were as follows:

- Learning Rate = 1.0 (Default Value)
- Momentum = 0.9 (Default Value)
- Input Noise = 0.0 (Default Value)
- Testing Tolerance = 0.3 (Default Value)
- Training Tolerance = 0.05 (Modified from Default Value of 0.1)
- Epochs per Update = 1.0

The default values were used in most cases except for the training tolerance. The reason was that the default values are positioned at a number that will generally give the best results. The training tolerance was modified to force the neural network numbers to come closer to the target numbers, which in turn forced the neural network values to approach the outputs that were generated from the targets.

The final phase of this project was to update the costs from CORA. This was accomplished in several ways. Most of the information came from the internet from two sources: the United States Environmental Protection Agency and the Ground-Water Technology Analysis Center. In other cases the cost data were collected from actual environmental design engineers in private practice. These updated costs allow other users of this neural network to have a base cost for twenty types of remediation schemes for both containment and treatment remediation. As with CORA generated cost projections, the costs are subject to change over time and with actual site conditions.

This project was able to train the neural network with a slightly dated expert system. This trained neural network is now ready to be used in place of the original expert system to precisely suggest remediation alternatives for VOCs in a homogeneous contaminated saturated soil. This network is fully updatable as cost or the technologies within CORA or subsequent neural network model change. Restructuring the spreadsheet

that is contained within the neural network is readily accomplished. As such this neural network is configured to allow future users to update it with different types of chemicals, new and innovative technologies, and cost. The basis of the information for the updates will either come from CORA or from other sources that the user obtains from either professional journals or the engineering trade. With additional work, this neural network could have the capability of predicting any type of chemical contaminant for which data exists. This research project has proven to be a success and should be further explored to help the environmental engineering field.

REFERENCES

- Amean, B. (1998), Landfill Designer, BFI Landfill, Phone interview, Tulsa, OK: (March 31).
- Aziz, A., Rashid, A. and Wong, K. V. (1992), "A Neural Network Approach to the Determination of Aquifer Parameters," *Ground Water*, (March-April), Vol. 30, No. 2. pp. 164-166.
- Berry, D. and Hart, A. (1990), *Expert Systems Human Issues*, New York, NY: Chapman and Hill.
- Bielawski, L. and Lewand R. (1988), *Expert Systems Development Building PC-Based Applications*, Wellesley, MA: QED Information Sciences, Inc.
- Castillo, E., Gutierrez, J. and Hadi, A. (1997), *Expert Systems and Probabilistic Network Models*, New York, NY: Springer-Verlag New York, Inc.
- CH2M Hill, (1990), *Cost of Remedial Action User Manual Version 3.0*, Washington D.C.: Environmental Protection Agency.
- Environmental Protection Agency (1998), <http://clu-in.com>.
- Findley, R. W. and Farber, D. A. (1991), *Environmental Law in a Nut Shell*, St. Paul, MN: West Publishing Company.
- Gifford, M., Gifford, S. and Gifford, M. (1997), "Net Results," *Civil Engineering*, (January), pp. 58-60.
- Gillies, A. C. (1991), *The Integration of Expert Systems Into Mainstream Software*, New York, NY: Chapman and Hall Computing.
- Ground-Water Remediation Technology Analysis Center (1998), <http://www.gwrtac.org>.
- Masters, G. M. (1991), *Introduction to Environmental Engineering and Science*, Englewood Cliffs, NJ: Prentice Hall.

- Powers, M. (1994), "Beating the Regulatory Odds with New Simulation Models," *Engineering News Record*, (June 20), Vol. 232, pp. 20-23.
- Prerau, D. S. (1990), *Developing and Managing Expert Systems*, Reading, MA: Addison-Wesley Publishing Company.
- Price, L. (1998), Public Works Director, City of Sand Springs, Phone interview, Sand Springs, OK: (March 31).
- Raeth, P. G. Ed., Williams C. (1990), "Expert Systems, Knowledge Engineering, and AI Tools-An Overview," *Expert Systems-A Software Methodology for Modern Applications*, Los Alamitos, CA: IEEE Computer Society Press.
- Rogers, L., Dowla, F. and Johnson V. (1995), "Optimal Field-Scale Groundwater Remediation Using Neural Networks and the Genetic Algorithm," *Environmental Science and Technology*, Vol. 29, No. 5, pp. 1145-1155.
- Shih, Y. (1994), *Neuralyst User's Guide (v1.4)*, Pasadena, CA: Cheshire Engineering Corporation.
- Siegel, P. (1986), *Expert Systems a Non-Programmer's Guide to Development and Applications*, Blue Ridge Summit, PA: TAB Professional and Reference Books.
- Sullivan, T. F.P., Ed. (1995), *Environmental Law Handbook Thirteen Edition*, Rockville, MD: Government Institutes, Inc.
- Taner, I., and Schultze, P. (1998), Environmental Engineers, A & M Engineering, Phone interview, Tulsa, OK: (March 31).
- Walker, T. C. and Miller R. K. (1986), *Expert Systems 1986*, Madison, GA: SEAI Technical Publications.
- Wilson, R. and Sharda, R. (1992), "Neural Networks: 'Brain Metaphors' of Information Processing Turns Heads in OR/MS Community," *OR/MS Today*, (August), pp. 36-42.

Figures 6 and 10.

RUN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Select all waste types that apply to the site in question	
Q1 HCUS	
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1

Select a response action that you wish to consider for the appropriate waste type	
Q2	
CONT	TREAT
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Hydraulic Conductivity - Select what best describes the soils	
Q3	
CLAY	SILT
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Select all types of contaminants in the soil	
Q4 VOC	
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1

Excavation of contaminated soils will result in acceptable environmental or public impacts	
Q5	
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1

Onsite incineration is precluded based on space (2-12 acre minimum clean space) and/or locational considerations	
Q5-a	
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1

Contaminated unsaturated soil in question for landfills is a hazardous waste	
Q6	Q7
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1

Contaminated unsaturated soil concentration is likely to be above the land disposal restrictions	
Q6	Q7
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1

Onsite RCRA landfill for contaminated unsaturated soil is reasonable based on available clean space, landuse, and	
Q8	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Select all types of contaminants in the leachate from landfill	
Q8-a VOCs	
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1

Contaminated unsaturated soils landfill footprint is in a 100-year flood plain	
Q8-b	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Shallow aquifer which would preclude a below grade landfill for contaminated unsaturated soil	
Q9	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Discharge option - Water reflection	
Q10	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Discharge Option - water infiltration	
Q10-a	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Discharge Option - discharge to POTW	
Q10-b	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Discharge Option - discharge to surface water	
Q10-c	
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1

Hydraulic conductivity of saturated zone	
Q10-d	
SILT	CLAY
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Water table greater than 5 feet below surface	
Q10-e	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Ash is a hazardous waste	
Q10-f	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Concentration of contaminant in ash is above land disposal restrictions	
Q10-g	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Onsite RCRA landfill for solidified ash is reasonable based on available clean space, landuse, and locational considerations	
Q10-h	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Types of contamination in the leachate from landfill	
Q10-i VOC	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0

Shallow aquifer would preclude a below grade landfill for solidified ash	
Q10-j	
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0

VITA

Matthew R. Thompson

Candidate for the Degree of

Master of Science

Thesis: USING A NEURAL NETWORK TO UPDATE AN EXPERT SYSTEM

Major Field: Environmental Engineering

Biographical:

Personal Data: Born in Stillwater, Oklahoma, On August 29, 1973, the son of Larry and Celia Thompson.

Education: Graduated from Charles Page High School, Sand Springs, Oklahoma in May 1991; received Bachelor of Science degree in Civil and Environmental Engineering from Oklahoma State University, Stillwater, Oklahoma in December 1996. Completed the requirements for the Master of Science degree with a major in Environmental Engineering at Oklahoma State University in May, 1998.

Experience: I have worked in the West Texas Oil fields for two straight summers (1996-1997). The first summer (1996) I worked in the environmental engineering department working with SARA Title III and NORM studies. The second summer (1997) I worked in the Heavy Oil Team performing reservoir engineering. This consisted mainly of performing steam flood efficiencies. I am currently employed by the Civil and Environmental Engineering department at Oklahoma State University as a teaching assistant for surveying. I have taught this surveying class since the spring of 1996 to the present.

Professional Memberships: National Society of Professional Engineers, Oklahoma Society of Professional Engineers, American Society of Civil Engineers.